# Physicians versus Midwives: Returns to Childbirth Technologies for Low-Risk Births*

N. Meltem Daysal
University of Southern Denmark and IZA

Mircea Trandafir
University of Southern Denmark

Reyn van Ewijk
University Medical Centre Mainz
and University of Mainz

This Version: December 2013

**Abstract**

We investigate the impact of obstetrician supervision, as opposed to midwife supervision, on the health of low-risk newborns. We exploit a unique policy rule in the Netherlands that creates a large discontinuity in the probability of a low-risk birth being attended by an obstetrician at gestational week 37. Using a fuzzy regression discontinuity design, we consistently find no health benefits from obstetrician supervision, despite increased neonatal intensive care unit admission rates among births supervised by obstetricians. These results indicate potential cost savings from shifting supervision of low-risk deliveries, which represent the vast majority of all births, from obstetricians to midwives.

Keywords: Medical technology, birth, midwife, mortality
JEL Classifications: I11, I12, I18, J13

---

*Daysal: University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: meltem.daysal@sam.sdu.dk); Trandafir: University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: mircea.trandafir@sam.sdu.dk); Van Ewijk: IMBEI, University Medical Centre Mainz, Obere Zahlbacher Str. 69, 55131 Mainz, Germany (email: vanewijk@uni-mainz.de). Abby Alpert, John Cawley, Yinging Dong, Amanda Kowalski, Martin Salm, Diane Whitmore Schanzenbach, Emilia Simeonova, Kosali Simon, Arthur van Soest, and participants at Aarhus University, Cornell University, Impaq International, University of Mainz, University of Munich, University of Southern Denmark, Universite de Sherbrooke, Tilburg University, European Economic Association Meetings and EuroEpi Meetings provided helpful comments and discussions. We thank Perinatale Registratie Nederland (PRN) for making the data available and PRN staff, in particular Chantal Hukkelhoven, for assistance with the data. Tjeerd van Campen and Iris van Dam provided able research assistance. The authors bear sole responsibility for the content of this paper.

# 1 Introduction

Medical expenditures increased tremendously over the last few decades throughout the entire developed world. In the United States, the share of health care expenditures in the GDP more than tripled between 1960 and 2009 (OECD, 2012). Most of this increase is attributed to changes in medical technologies (Newhouse, 1992). However, there is substantial heterogeneity in treatment effects across patients, and thus there are some technologies that are highly effective for some populations while having little benefits for others (Chandra and Skinner, 2012). Therefore, decision-makers could slow down the growth of medical expenditures if they can reduce the use of inefficient technologies. One such strategy relies on the expanded use of lower-cost physician extenders instead of higher-cost physicians for patients who benefit little from being treated by the latter (Fuchs, 1998).

In this paper, we examine the effects of increased use of physician extenders in the context of childbirth technologies. In particular, we investigate the impact of obstetrician/gynecologist, as opposed to midwife, supervision of deliveries on the health (7-day and 28-day mortality and Apgar score) of low-risk newborns.[1] Focusing on childbirth technologies is especially important given the substantial increase in spending for the very young over the past decades (Cutler and Meara, 1998), raising concerns over the efficient utilization of health care resources. In this context, the use of midwives for the supervision of low-risk births is a cost-saving measure rooted in the idea that these deliveries should not lead to situations requiring medical interventions. However, this can lead to worse newborn health outcomes in case complications occur during deliveries due to delays if there is no obstetrician in close proximity, or if the OB/GYN cannot obtain all the relevant medical information because of the urgency of the intervention.

[1] Gynecology and obstetrics are overlapping medical specializations that deal with the female reproductive organs in their non-pregnant and pregnant state, respectively. As all gynecologists are generally also obstetricians, physicians attending pregnancies and deliveries are often called "OB/GYN," for obstetrician/gynecologist. In the remainder of the text we use OB/GYN and obstetrician interchangeably.

Empirical estimation of the returns to obstetrician supervision is complicated by selection issues: women with worse expected birth outcomes usually give birth under the supervision of an OB/GYN, leading to biased estimates in simple regressions. In order to eliminate this bias, we exploit a policy rule in the Netherlands that provides an exogenous source of variation in the type of medical professional supervising a low-risk birth.

The Dutch system is unique in its division between the primary care provided by midwives and the secondary care provided by obstetricians. Low-risk women (women without known medical risk factors) start their pregnancy under the supervision of a midwife and stay under the supervision of a midwife as long as no additional risk factors appear. Midwives, who are prohibited by law from performing any medical intervention, also supervise the delivery and no OB/GYN is present. However, if labor is premature (before 37 completed gestational weeks), the woman should be referred to an obstetrician who will supervise the delivery. This "week-37 rule" generates a discontinuity in the probability of being treated by an OB/GYN at gestational week 37 among low-risk women, motivating the use of a regression discontinuity (RD) design. It is worth noting that the rate of planned Caesarean sections is generally very low in the Netherlands and that planned Caesarean sections do not occur among low-risk women.[2]

A valid RD design needs to satisfy two conditions. First, it has to be based on a policy relying on a sharp and arbitrary cutoff. Second, the measure used in implementing the policy should not be under the precise control of the targeted individuals. If these conditions are satisfied, then random variation around the cutoff will partly determine when the policy is implemented (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). The week-37 cutoff provides an ideal case for an RD design. The medical literature acknowledges the arbitrariness of the threshold of 37 completed gestational weeks used to

[2]Only around 7 percent of all births are primary Caesarean sections (i.e., planned before the start of delivery). Most of these are for medical reasons and among women not classified as low-risk. Elective Caesarean sections for non-medical reasons are very rare and virtually non-existant around the 37-week cutoff. As detailed later in the paper, all planned Caesarean sections are excluded from our analysis sample.

define prematurity (e.g., Kramer et al., 2012). Neither the mother nor the fetus experience any sharp changes in their health risk between gestational days 258 and 259. Moreover, the same 37-week threshold is used throughout the world regardless of how gestational age is measured — from the first day of the last menstrual period or by means of an ultrasound, although the former is known to overestimate true gestational age. Finally, it is not possible for women having a spontaneous birth to precisely control the timing of their birth as there are no medical tests that can accurately predict prematurity (Institute of Medicine, 2007).

Using data on the universe of births in the Netherlands between 2000–2008, we show that the week-37 rule generates substantial variation in the medical professional supervising the birth. In particular, we find that the probability that a spontaneous low-risk birth is supervised by an obstetrician increases by almost 60 percentage points below the 37-week threshold. We confirm that the variation in OB/GYN supervision generated by the week-37 rule is as good as random in two ways. First, we show that there are no heaps in the frequency of births around the cutoff. Second, we show that the distribution of a wide range of covariates is generally smooth around the week-37 cutoff.

We proceed to estimate the causal effect of obstetrician supervision on infant health outcomes by using the variation induced by the week-37 rule in an instrumental variable (IV) framework. Despite the substantial variation in OB/GYN supervision, our results indicate that average newborn health outcomes are remarkably similar across the week-37 cutoff. Our IV estimates of the OB/GYN effect are consistently insignificant and of the wrong sign, indicating no health benefits from obstetrician supervision. These results are robust to a host of checks, such as the inclusion of a full set of covariates, donut regressions that exclude observations with gestational ages of 258 and 259 days, different polynomial degrees in the running variable, alternative bandwidths, various definitions of the running variable, and non-linear specifications.

Our estimates identify a combination of two effects: first, the effect of the OB/GYN supervising the delivery; and second, potential differences in the use of medical technology between OB/GYN supervised and midwife supervised

deliveries. Previous research indicates that OB/GYN supervised deliveries are more likely to end up with medical interventions (e.g., see Sandall et al., 2013). Consistent with this research, we find that admissions to neonatal intensive care units (NICU) in the first 7 days of life occur more frequently among deliveries supervised by an OB/GYN.[3] Given that 7-day and 28-day mortality are similar among births supervised by midwives and by OB/GYNs, this result underlines the potential for cost-saving by allowing midwives to supervise low-risk deliveries.

Our estimation strategy yields the local average treatment effect (LATE) for low-risk births close to 37 completed gestational weeks that were supervised by an OB/GYN only because of prematurity. This may not reflect the average treatment effect of obstetrician supervision, but the compliers in our setting comprise almost 60 percent of the sample and have similar observable characteristics to the overall analysis sample of low-risk births. Given that infant health is generally an increasing function of gestational age (up to around 42 completed gestational weeks), our results cannot be generalized to gestational ages below those studied in this paper but they are likely to apply to low-risk at-term births with higher gestational ages.

Our study adds to the handful of economic studies that provide convincing evidence on the returns to childbirth technologies. A large part of this literature investigates treatments for high-risk newborns, such as those with very low birth weight (Almond et al., 2010; Bharadwaj et al., forthcoming;

[3]While we do not have data on other treatments, other studies report that many hospitals in the Netherlands regularly admit prematurely-born children for observation and some hospitals give antibiotics to women whose water breaks before week 37 in order to reduce the risk of infection (Schakel and Bekhof, 2010). The impact of these treatments is included in our estimated effect of OB/GYN supervision. To the extent that these medical treatments have non-negative effects on newborn health, they will bias our results towards overestimating the benefits of OB/GYN supervision. Finally, deliveries supervised by obstetricians and midwives may also differ in terms of place of delivery. Obstetricians can only supervise births in a hospital setting. Midwives are allowed to supervise home deliveries, but not before week 37. As a result, the share of home births is much higher to the right of the week-37 cutoff. Daysal et al. (2012) show that home births in the Netherlands lead to higher infant mortality, particularly among women living in low-income neighborhoods. This again suggests that any bias in our results would lead to an overestimate of the benefits of OB/GYN supervision.

Cutler and Meara, 2000; Freedman, 2012). In contrast, we focus on *low-risk* newborns which constitute the vast majority of births. In addition, we investigate the specific role of physician extenders, which is largely unexplored in the economic literature.[4] Our paper also adds to the existing medical literature on returns to midwifery care. As we detail in section 2.2, these studies generally rely on simple regression models comparing outcomes among subsamples of low-risk women who give birth under the supervision of a midwife or an obstetrician, after controlling for observable differences. The major drawback of this approach is a potential selection bias due to the endogeneity in provider choice. The few studies that rely on randomized controlled trials, on the other hand, suffer from small sample sizes and thus cannot focus on rare outcomes such as infant mortality.

Our results point to potential cost savings from increased use of midwifery care for low-risk deliveries. These findings are relevant to the ongoing policy debates on cost reduction through increased used of physician extenders. A growing number of women in developed countries are giving birth with a midwife. For example, 7.6% of all births in the US are attended by a midwife and this share is rising (Kochanek et al., 2012). 7.6% of UK deliveries are attended by midwives in midwifery units or at home, where no OB/GYNs are immediately available (Redshaw et al., 2011). In addition, there are multiple calls for an increased role for midwifery care in medical journals, policy briefs and in the popular press. For example, a recent report by the Institute of Medicine (2011) recommends the expansion of duties for physician extenders, including those trained to supervise births. Under these circumstances, understanding the impact of physician extenders on health outcomes and on cost containment is likely to remain an important policy topic in the coming years.

The remainder of the paper is organized as follows. In section 2, we present the Dutch obstetric care system and the week-37 rule, as well as a brief summary of the relevant literature. The data used is described in section 3, while

[4]One exception is Miller (2006) who uses variation in mandated insurance coverage of midwifery services to estimate the effects of midwifery-promoting public policies on maternal and newborn health outcomes. However, the results of this study may still suffer from a bias due to the selection of women into midwifery care.

section 4 outlines the empirical framework. The results are presented in section 5 along with a discussion of the validity of our RD design and robustness checks. Section 6 concludes.

# 2 Background

## 2.1 The Dutch obstetric system

The Dutch obstetric system is characterized by a strict role division between midwives and obstetricians/gynecologists, in which midwives play a larger role than in many other developed countries. This system resulted from first, a philosophy that pregnancy and delivery are natural processes that do not require attendance by an OB/GYN, as long as there are no deviations from the perfectly normal course; and second, from cost-reducing measures that led to a set of rules limiting which pregnancies and deliveries should be supervised by OB/GYNs.[5] These rules were first implemented in 1958 and over time resulted in the "List of Obstetric Indications" (LOI) prescribing the conditions that can lead to a referral from midwife to OB/GYN (Amelink-Verburg and Buitendijk, 2010).

Pregnancies in the Netherlands start under supervision of a midwife as long as none of the conditions described in the LOI is present. If at least one such condition is present, the pregnancy and the birth should be supervised by an OB/GYN. If a listed condition arises during pregnancy, a referral needs to be made at that point in time. The LOI contains four types of criteria that lead to a referral: non-gynecological pre-existing conditions (e.g., diabetes, alcoholism or psychiatric disorders); gynecological pre-existing conditions; obstetric anamnesis (cesarean section, very premature births or severe complications during previous deliveries); and conditions arising or first diagnosed during pregnancy (e.g., hyperemesis gravidarum, infections, plurality, gestational hypertension, or blood loss) (CVZ, 2003). Referrals for reasons

[5]This philosophy and these rules also explain why planned Caesarean sections are rare in the Netherlands compared to most other developed countries.

not listed in the LOI are not allowed and physician fees are not covered by insurance plans in such cases. Also, women are not allowed to directly contact an OB/GYN.

This risk selection system divides pregnant women into two groups. High-risk women are those referred to an OB/GYN at any point during pregnancy (before the onset of labor). Their prenatal care is provided by the OB/GYN from the moment of the referral and they are required to give birth in a hospital under the supervision of an OB/GYN. Low-risk women are those who do not develop any complications during pregnancy and are under the care of a midwife at the onset of labor. Their prenatal care is provided entirely by midwives and their deliveries are supervised by a midwife with no OB/GYN present unless a complication arises during delivery. However, if the onset of labor occurs before 37 completed gestational weeks, the LOI prescribes that the woman should be referred to an obstetrician. For the purpose of this rule, gestational age is measured in full days from the last menstrual period. This "week-37" policy rule generates plausibly exogenous variation in births attended by different providers and is the basis of our empirical strategy.

## 2.2 Previous Literature

This paper fits broadly in the previous economics research on returns to medical technologies. A large part of this literature investigates treatments for heart attack patients (Cutler et al., 1998; McClellan and Newhouse, 1997; McClellan and Noguchi, 1998; Skinner et al., 2006). More recently, a growing number of papers examine returns to childbirth technologies, with a special focus on treatments for (very) low birth weight children. Increased treatments for this group were generally shown to reduce mortality (Almond et al., 2010; Bharadwaj et al., forthcoming; Cutler and Meara, 2000), with the exception of Freedman (2012), who finds no health gains from the deregionalization of neonatal intensive care units.

Research on the returns to medical technologies for low-risk infants is limited. Currie and MacLeod (2008) find that tort reforms increased the use of

C-sections and led to lower rates of preventable complications but did not have any effects on Apgar score. Almond and Doyle (2011) show that longer hospital stays do not affect health outcomes after uncomplicated deliveries. On the other hand, Daysal et al. (2012) find that giving birth in a hospital (as opposed to home) leads to reductions in the mortality of low-risk newborns. Most relevant to the current study is the paper by Miller (2006) examining the impact of midwifery-promoting policies on maternal and infant health outcomes. This study uses state-level variation in mandated insurance coverage of midwifery service and finds that midwifery-promoting public policies had no significant effect on maternal mortality or Apgar scores, but were associated with lower neonatal mortality. However, this econometric strategy cannot eliminate the potential selection bias due to women choosing a certain provider type (or being assigned to one by their insurance provider) based on unobservable characteristics.

In the medical field, several studies investigate the effects of midwife versus OB/GYN supervision. A few of these are randomized controlled trials.[6] They have relatively small sample sizes, which preclude them from studying effects on rare outcomes such as mortality. Instead, these papers focus on rates of medical interventions such as labor induction and cesarean sections (which are ultimately performed by physicians). Intervention rates are found to be lower after midwife-supervised deliveries (Chambliss et al., 1992; Harvey et al., 2007; Rosenblatt et al., 1997; Tumbull et al., 1996). Other studies compare women *choosing* different provider types while controlling for observable characteristics. Most of these find equal or lower mortality rates for midwife-supervised deliveries and again lower rates of medical interventions (Birthplace in England Collaborative Group, 2011; Janssen et al., 2002; MacDorman and Singh, 1998). All these observational studies are likely biased as the choice of birth attendant is endogenous.[7] For example, MacDorman and Singh (1998) report

[6]For a review of this literature, see Sandall et al. (2013).

[7]One exception is a recent study by Evers et al. (2010) who find that delivery-related infant mortality in the Netherlands is higher among low-risk births supervised by a midwife than among high-risk births supervised by an OB/GYN. This study and the interpretation of its findings have been hotly debated in the Netherlands and the study was heavily criticized

that the risk of delivering a low birth weight infant is 31 percent lower "as a result of" midwife supervision, although birth weight is obviously unlikely to change due to the medical professional supervising the birth. Our study explicitly corrects for the endogeneity of birth attendant exploiting the exogenous variation induced by the week-37 rule.

# 3 Data

Our primary analysis uses data from the Perinatal Registry of the Netherlands (Perinatale Registratie Nederland, PRN) for the years 2000–2008. PRN is an annual dataset that covers approximately 99 percent of the primary care and 100 percent of the secondary care provided during pregnancy and delivery in the Netherlands (de Jonge et al., 2009). It is constructed by linking individual birth records provided separately by midwifes (LVR-1), obstetricians/gynecologists (LVR-2) and paediatricians (LNR).[8] The data includes detailed information on the birth process. For each delivery, we observe the date and time of birth, type of birth attendant (midwife or OB/GYN), delivery location, method of delivery (natural birth, planned C-section, emergency C-section, labor augmentation, induction, use of forceps and vacuum etc.) as well as the presence of complications during pregnancy or delivery. In the case of complications, we can observe if the woman was referred from a midwife to an OB/GYN and, if so, the date and reason of referral. PRN also includes a number of variables pertaining to short term infant health outcomes (including mortality and the Apgar score) and limited information on diagnosis and treatment (e.g., NICU admission within the first 7 days of life, emergency C-section). Finally, it provides rich background information on newborns (gender, gestational age in days, birth weight, parity and plurality) and basic demographic characteristics of mothers (age, ethnicity, residential postal code).

by some researchers in the medical literature (e.g., de Jonge et al., 2010).

[8] PRN data does not include information on births supervised by general practitioners which constitute a very small share of all primary care deliveries (Amelink-Verburg and Buitendijk, 2010).

Some of our analyses complement the individual-level PRN data with a secondary postal code-level data set from Statistics Netherlands (Kerncijfers postcodegebieden 2004). These data provide a snapshot of characteristics in the postal code of residence of the mother[9] as of January 1, 2004 and include information on the average household income, the average area density[10] and the share of 0-15 year-olds in the postal code.

We construct our analysis sample in three steps. The full sample includes data on 1,630,062 newborns. First, we exclude observations for which the type of birth attendant and gestational age are missing. Second, we exclude stillbirths and cases in which gestational age might be manipulated (planned C-sections and induced and stimulated births). Third, we restrict our sample to low-risk mothers because the week-37 rule applies only to them. As discussed in section 2.1, low-risk mothers are those under the care of a midwife at the onset of labor. However, some low-risk women may be referred to an obstetrician shortly before delivery due to an impending premature delivery, usually because contractions started or water broke. In addition, there is some discretion in how medical professionals classify a referral. For example, a woman referred during the contraction phase but before the pressing phase might be coded as a referral during delivery by some and as a referral before delivery (i.e., during pregnancy) by others. In the latter case, these women are coded as "high risk" in our data since they are under the supervision of an OB/GYN at the onset of labor. If such referrals are not random, women coded as "low-risk" would constitute a selected sample. For that reason, we also include in our sample deliveries that started under the supervision of an OB/GYN but for which the referral occurred at most one day before delivery.[11] Finally, we focus on observations with gestational age within 14 days of

[9]Postal codes in the Netherlands have 6 digits and our data includes 4-digit postal codes for mothers. It should be noted that postal codes in the Netherlands are much smaller than zip codes in the United States. The average 4-digit area has 4,075 inhabitants and a land surface of 8.5 square kilometers (3.28 square miles).

[10]This is the average number of addresses per square kilometer in a circle with a radius of 1 km around each address in the postal code.

[11]Date of referral is missing for a substantial number of observations within our bandwidth (76,344 deliveries). In section 5.4, we check the sensitivity of our results to the exclusion of

the 37-week cutoff (day 258/259). This leaves us with an analysis sample of 150,471 newborns.[12]

We focus on three outcome variables that capture the short term health of newborns: 7-day mortality, 28-day mortality and low Apgar score.[13] Our main explanatory variable is an indicator for OB/GYN versus midwife supervision of delivery. Prenatal care for all women in our sample is provided by midwives, so any differences in postnatal outcomes between deliveries supervised by a midwife and those supervised by an OB/GYN should be due to the medical attendant. However, one complication is that our data does not perfectly enable us to distinguish between situations where the obstetrician supervised a birth from the start of delivery and situations where the OB/GYN was called later on because of complications during the delivery. Therefore, we construct three measures of OB/GYN supervision that are affected in different ways by these two scenarios and we compare the results across these three measures. A consistently estimated effect of OB/GYN supervision across these measures would indicate the "true" effect of obstetrician supervision on newborn outcomes.

The first measure is an indicator that equals 1 for deliveries coded in our data as supervised by an OB/GYN from the onset of labor. As we restricted our sample to women under the care of a midwife until one day before delivery, this measure captures those women who were referred very shortly before delivery and for whom the referral was coded as occurring during pregnancy. Among these deliveries, we expect that a majority of the deliveries with gestational age less than 37 completed weeks are referred to obstetricians because of prematurity.

However, as mentioned above, some of the referrals for prematurity might

these observations.

[12] We discuss bandwidth selection in section 4.2. In section 5.4, we show that our results are robust to alternative sample restrictions.

[13] We do not have information on longer term mortality rates. Apgar is measured 5 minutes after birth and summarizes the health of newborns based on five criteria: appearance (skin color), pulse (heart rate), grimace response ("reflex irritability"), activity (muscle tone), and respiration (breathing rate and effort). The score ranges from 0 to 10 with higher scores indicating better health. A low Apgar score indicates an Apgar score of less than 7.

be coded as having occurred "during delivery." In order to take these deliveries into account, we construct a second measure of OB/GYN supervision that adds to the first measure women coded as referred during delivery and for whom prematurity is included among the recorded reasons for referral.[14] Since it is likely that women who are referred to an OB/GYN because of prematurity are supervised by an obstetrician from the onset of labor, this measure captures a greater share of those who started delivery with an OB/GYN due to prematurity than the first measure.

Finally, it is known that the recording of reasons for referral is sometimes incomplete. Our third measure therefore classifies women as supervised by an obstetrician if an OB/GYN was present at any point during the delivery. This means that all women referred to an OB/GYN because of prematurity are now classified as supervised by an obstetrician.[15] Figure A2 in the Appendix describes the three measures of OB/GYN supervision.

A second explanatory variable crucial to our identification strategy is gestational age, measured as the number of days between the date of the mother's last menstrual period and the date of birth. Finally, some of our robustness checks include additional covariates, which can be classified into four groups. The first group (time effects) includes fixed effects for the year, month and day of the week of the birth. The second group (maternal characteristics) includes mother's age and ethnicity.[16] The third group (infant characteristics) includes birth weight and indicators for gender, parity, plurality, congenital anomalies and birth position.[17] The final group (postal code characteristics) includes the

[14]Midwives and OB/GYNs each record three reasons for referral. We classify a referral as due to prematurity if at least one of the four codes assigned to prematurity was among these recorded reasons for referral.

[15]All referrals due to complications during delivery are also classified as supervised by an OB/GYN under this measure. This should not influence our results since our estimates are driven by referrals due to prematurity (in instrumental variables terminology, referrals due to complications are always takers).

[16]We include indicators for six maternal age categories (less than 20, 20–24, 25–29, 30–34, 35–39, 40 and above) and three maternal ethnicity categories: Dutch, Mediterranean and others (Moroccans and Turks, commonly identified as "Mediterraneans," represent the majority of the immigrant population in the Netherlands).

[17]Specifically, we include birth weight in grams and indicators for very low birth weight (less than 1,500 grams), low birth weight (between 1,500 and 2,500 grams), gender, parity

average characteristics of the postal code of residence of the mother: household income, area density and the share of 0–15 year-olds.[18]

# 4 Empirical Strategy

## 4.1 Empirical Framework

The goal of this paper is to estimate the effect of physician supervision of births, as opposed to midwife supervision, on infant health outcomes. Simple regressions are likely to provide biased estimates because of the potential endogeneity of the attending medical professional. In particular, the concern is that OB/GYNs supervise births with worse observed and potentially unobserved health characteristics. This is especially the case in the Netherlands due to risk selection, as detailed in section 2.1.[19] In order to overcome this endogeneity, we use the exogenous variation provided by the week-37 rule. According to this rule, births occurring before 37 completed gestational weeks should be supervised by an OB/GYN in a hospital. This results in a discontinuity in the probability of supervision by an OB/GYN, suggesting a regression discontinuity (RD) design.

An RD design relies on the idea that if a policy requires a sharp and arbitrary cutoff for implementation and is based on a measure that is not perfectly controlled by the targeted individuals, then random variation around the cutoff will partly determine when the policy is implemented (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). The week-37 cutoff provides an ideal case for an RD design. It is based on an arbitrary threshold in the sense that there are no specific developmental changes that occur in the

(first born), plurality, congenital anomalies (mild and severe) and birth position (breech birth and other).

[18] Some of the control variables (newborn gender, birth weight and parity, mother's age, and postal code characteristics) are missing for a very small number of observations (less than 0.03 percent for individual characteristics and less than 0.8 percent for postal code characteristics). We replace these missing values with the sample average of the corresponding variable and we include as additional controls indicators for missing values for each variable.

[19] Note again that women are not allowed to choose themselves whether a midwife or an OB/GYN supervises their deliveries.

fetus or in the mother between day 258 and day 259. Kramer et al. (2012, p.111) note that "[i]nfants born before 20 weeks or at 37 or 38 weeks share many features with births at 20–36 weeks, including etiological and prognostic features," and thus conclude that the choice for the upper (37 weeks) and lower (20 or 22 weeks) bounds for defining a preterm birth are arbitrary. The arbitrariness of the week-37 cutoff is further supported by the fact that the definition of prematurity uses the same threshold regardless of whether gestational age is measured since the last menstrual period or by means of an ultrasound, although the former overestimates gestational age on average. In addition, a recent report by the Institute of Medicine (2007, p.3) notes that "[t]o date, no single test or sequence of assessment measures that may accurately predict preterm birth are available." This suggests that, in our sample of spontaneous births, expectant mothers cannot precisely manipulate the timing of their birth so as to control their assignment to different medical professionals. These two points indicate that the variation in OB/GYN-supervision near the week-37 cutoff is as good as random.

It is worth mentioning that crossing the week-37 threshold can change both the medical professional supervising the birth as well as the set of medical treatments applied to the newborn. Preterm births are likely to receive different (additional or alternative) treatments than births occurring after 37 completed gestational weeks. Therefore, the treatment effect identified by the week-37 rule is a combination of the effect of OB/GYN supervision and that of medical technologies. To the extent that these medical technologies do not reduce the chances of survival of preterm newborns (Cutler and Meara, 2000; Almond et al., 2011; Freedman, 2012; Almond and Doyle, 2011; Daysal et al., 2012), this implies that our results overestimate the health benefits of OB/GYN supervision.

In our framework, the probability of OB/GYN supervision does not "jump" from 1 to 0 using either of our classifications when gestational age increases from just under to just over 37 weeks for two reasons. First, the week-37 rule is not perfectly enforced, meaning that not all the infants born before 37 com-

pleted gestational weeks are supervised by an OB/GYN.[20] Second, low-risk women who are under the care of a midwife at the onset of delivery can be referred to an OB/GYN for reasons other than prematurity, including complications arising during delivery, slow progression, or the need for pain relief medication. As a result, some of the births with at least 37 completed gestational weeks are supervised by OB/GYNs. Hahn et al. (2001) show that the estimation of causal effects in such a "fuzzy" regression discontinuity framework is numerically equivalent to an instrumental variable (IV) approach within a small interval around the discontinuity. We provide details on the selection of this bandwidth in the next section.

Our empirical strategy can be described by the following local-linear regressions:

$$Y_{iat} = \beta_0 + \beta_1 OB/GYN_{iat} + \beta_2(a - 258) + \beta_3 W37_a(a - 258) + \epsilon_{iat}, \quad (1)$$

$$Y_{iat} = \delta_0 + \delta_1 W37_a + \delta_2(a - 258) + \delta_3 W37_a(a - 258) + v_{iat}, \quad (2)$$

$$OB/GYN_{iat} = \alpha_0 + \alpha_1 W37_a + \alpha_2(a - 258) + \alpha_3 W37_a(a - 258) + u_{iat}, \quad (3)$$

where the unit of observation is infant $i$ born in year $t$ at gestational age $a$. The first equation represents the structural model relating an infant health outcome $Y$ to the main variable of interest, one of our three indicators for OB/GYN supervision, and a first-degree polynomial in our running variable (normalized gestational age) that is allowed to vary on both sides of the discontinuity.[21] In order to estimate the causal effect of OB/GYN supervision on infant health outcomes, we instrument for $OB/GYN_{iat}$ using an indicator for prematurity:

[20] The imperfect application of the rule does not invalidate the RD design unless the assignment of births to OB/GYNs or to midwives can be precisely manipulated around the discontinuity. In addition, adherence to the rule is likely not related to differential access to obstetric care as all pregnant women have insurance coverage and 98 percent of the Dutch population lives within a 30-minute drive from an obstetrics ward (Nationale Atlas Volksgezondheid, 2011). We discuss this aspect in more details in section 5.1.

[21] $W37_a$ is an indicator for gestational age of less than 37 completed weeks. We normalize the running variable to zero at 258 gestational days because the "treatment" we estimate is applied below the discontinuity. With this definition of the running variable, the coefficient of interest $\beta_1$ captures the change in the infant health outcome as gestational age moves from 259 days (exactly 37 completed weeks) to 258 days.

$W37_a = I(a < 259)$. Equation (2) represents the reduced form relationship between infant health outcomes and our instrument, in which the parameter $\delta_1$ can be interpreted as an intention-to-treat effect. Finally, the last equation is the first stage regression that can be used to verify that the week-37 rule impacts significantly the fraction of births supervised by an OB/GYN.

Our baseline regressions use a rectangular kernel which places the same weight on all observations. This is equivalent to estimating OLS regressions within the chosen bandwidth (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). Since the running variable is discrete, we cluster the standard errors in all regressions at the gestational day level (Lee and Card, 2008).

Our estimation strategy identifies the local average treatment effect (LATE) for "compliers" around the week-37 cutoff. Compliers comprise the subsample of births that are supervised by an OB/GYN only because they were premature, but that would be supervised by a midwife if they occurred after 37 completed gestational weeks. A LATE may not reflect the average treatment effect of obstetrician supervision, but as we show later the share of compliers in our study is high. Moreover, deliveries at 37 completed weeks can be considered among the riskiest (in gestational age terms) at-term births. The fact that we find midwife supervision to be safe among this group suggests that it is also safe for low-risk at-term births after week 37.

## 4.2 Bandwidth Selection

Estimation in an RD framework is conducted within a small interval around the discontinuity. Larger bandwidths increase the degree of precision of the estimates, but also increase the risk of bias. The literature generally suggests two methods for selecting the optimal bandwidth of this interval: a rule-of-thumb approach and a leave-one-out cross-validation procedure (Lee and Lemieux, 2010). For each health outcome, the optimal rule-of-thumb bandwidth is given by the formula:

$$h_{ROT} = k \left[ \frac{R\hat{\sigma}^2}{\sum_{i=1}^{n} (\hat{m}_i'')^2} \right]^{1/5},$$

where $k$ is a parameter that depends on the kernel choice (2.702 for the rectangular kernel), $R$ is the range of the running variable, $n$ is the sample size, and $\hat{m}''(\cdot)$ and $\hat{\sigma}$ are the curvature and standard error of the regression of the health outcome on a fourth-degree polynomial in (normalized) gestational age, respectively.

The second approach is based on the calculation of a cross-validation function. For each health outcome and for a given bandwidth $h$, the value of the cross-validation function is:

$$CV_Y(h) = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2,$$

where $\hat{Y}_i$ is the predicted value of the health outcome from a regression of $Y$ on a first-degree polynomial in (normalized) gestational age. These regressions use observations within the bandwidth $h$ of observation $i$ (only to the left of observation $i$ if $i$ is to the left of the cutoff, and only to the right of observation $i$ if $i$ is to the right of the cutoff), but excluding observation $i$ itself. The optimal bandwidth is then the $h$ that minimizes the value of the cross-validation function.

Appendix Table A1 lists the optimal bandwidths for our selected health outcomes based on both methods, calculated separately for observations on each side of the discontinuity and overall, and Appendix Figure A1 plots the cross-validation functions using observations on both sides of the discontinuity. Our baseline regressions use a bandwidth of 14 days to the left and right of gestational day 258. In section 5.4 we investigate the robustness of our results to the choice of bandwidth.

# 5 Results

## 5.1 Tests of the Validity of the Regression Discontinuity Design

The validity of an RD design rests on the assumption that individuals do not have precise control over the assignment variable. Since there are no medical tests which can accurately predict prematurity and our analysis sample consists of spontaneous births, the variation in obstetrician-supervision near the week-37 cutoff can be reasonably thought to be as good as random. However, the key identification assumption of the RD design could be violated if women (or midwives) strategically misreport the gestational age at birth in order to affect the medical professional attending the delivery.

In order to test this, we examine in Figure 1 the frequency of births by gestational age within a 4-week interval around the cutoff. A discontinuity in the density of births around the week-37 cutoff would suggest manipulation of the running variable and thus invalidate our RD design (McCrary, 2008). Not surprisingly, the number of births is increasing in gestational age, with the vast majority of births occurring after 39–40 completed gestational weeks. However, visually, there is no significant jump in the number of births between days 258, when the week-37 rule applies, and 259, when it does not. More formally, we estimate a local-linear regression similar to the reduced form (2), using the logarithm of the number of births at each gestational age as the dependent variable. While the results of this regression indicate a statistically significant jump in the frequency of births at the discontinuity, we find that there are 17 other jumps between two consecutive gestational days in the full analysis sample that are of the same relative magnitude or higher than the jump between days 258 and 259. This suggests that the statistically significant increase in the number of births at the cutoff is not an irregular heap but merely a result of the small number of observations in each gestational age bin.[22]

[22] In section 5.4 we show results from "donut" regressions which exclude observations with gestational age of 258 and 259 days, where incentives for manipulation might be strongest (Barreca et al., 2011).

Next, we check whether there are differences in observable characteristics across the week-37 cutoff. If the RD design is valid, then the observable characteristics should be locally balanced on both sides of the week-37 cutoff. Figure 2 presents the means of selected covariates by gestational age in a 4-week interval before and after the cutoff.[23] As the Figure shows, the distribution of the covariates is smooth around the discontinuity. In order to examine this issue more formally, we also provide in Table 1 the means of covariates on either side of the discontinuity within a 14-day bandwidth after controlling for gestational age. The last column of the table provides the p-values for the test of equality of the means, clustered at the gestational day level.[24] The results reported in Table 1 confirm the visual evidence in Figure 2: observations just below the week-37 cutoff are similar to those just above the week-37 cutoff in terms of maternal characteristics (age, ethnicity), the majority of newborn characteristics (plurality, low and very low birth weight, congenital anomalies), and the average characteristics in the postal code of the residence of the mother (density, share of 0-15 year olds). There is also no evidence that variables are more likely to have missing values on one side of the discontinuity, suggesting no differences in misreporting. It is worth noting that even in the few cases where we find statistically significant differences, the difference in the magnitudes is very small. For example, infants born before day 259 are on average only 48 grams lighter than those born after the cutoff and they are 0.08 percentage points more likely to have missing birth weight. Similarly, the difference in the average monthly income in the postal code of mothers of preterm and at-term infants is €10.[25]

[23]For visual clarity, here and in the rest of the paper, we group data in 4-day bins starting from the cutoff. Figures with data at the daily level are included in the Appendix.

[24]This analysis is equivalent to estimating a local-linear regression similar to the reduced form (2) using the covariates as the dependent variable, with the difference in means below and above the cutoff (i.e., columns 1 and 2) representing the coefficient estimate for $W37_a$ and the corresponding p-value indicated in column 3. Appendix Table A2 lists the remaining covariates.

[25]The small statistically significant jump in birth weight is not suprising because birth weight and gestational age are particularly related to each other. Almond et al. (2010) exploit the variation in medical inputs across the very low birth weight threshold to estimate the marginal returns to medical care and also find a statistically significant jump in

Overall, the analyses in this section indicate that there is no evidence of manipulation of the running variable around the week-37 cutoff. In addition, we find no systematic evidence of discontinuities in the observable characteristics of the newborns and their mothers. This lends support to the claim that the variation in obstetrician-supervision near the week-37 cutoff is as good as random.

## 5.2 The Discontinuity in Obstetrician Supervision

If the Dutch institutional rule governing the supervision of premature births is binding, then we should observe a discontinuity in the share of births attended by an obstetrician at 37 completed gestational weeks. To examine this, in Figure 3 we plot the fraction of births that are supervised by an obstetrician for each of our three measures for gestational ages within a 4-week interval around the cutoff. Visually, there is a substantial jump up at the week-37 cutoff for all three measures, with newborns below the cutoff having higher rates of obstetrician supervision. The fact that there is a significant jump in our first measure of OB/GYN supervision confirms our expectation that some low-risk women coded as "referred during pregnancy" are indeed referred to an obstetrician because of an impending premature birth.

Next, we examine whether the difference in obstetrician supervision below and above the week-37 cutoff is statistically significant by estimating equation (3) within our baseline bandwidth of 14 days around the cutoff. The first row of each Panel in Table 2 provides the estimated first stage relationship between the measure of OB/GYN supervision and the instrumental variable. Each column of the Table corresponds to a different newborn health outcome, with each cell representing a different regression. The results suggest that a premature birth is a strong predictor of whether the birth is supervised by an obstetrician or by a midwife. The coefficient estimates for the week-37 indicator are highly significant and they indicate that premature births are, on average, 20–59 percentage points more likely to be supervised by an

gestational age at the very low birth weight cutoff.

obstetrician than births with at least 37 completed gestational weeks. The F-statistics testing the statistical significance of the week-37 indicator are well above the rule-of-thumb value of 10. In conclusion, the evidence suggests that the Dutch institutional setup provides significant variation in obstetrician supervision of newborns.

## 5.3 Baseline Results

In this section we present our baseline estimates of the effects of obstetrician supervision of births on newborn health outcomes. We start by analyzing the reduced form relationship between infant health outcomes and prematurity. Figure 4 plots the evolution of our three measures of newborn health as a function of gestational age within a 4-week window around the cutoff. The Figure indicates a smooth evolution of our three health measures across the week-37 cutoff, suggesting no significant health differences between births of slightly less and slightly more than 258 completed gestational days.

The lack of visual evidence of an effect is confirmed by our coefficient estimates from the reduced form equation, presented in the second row of each Panel in Table 2. The results, although imprecise, indicate no significant health differences between preterm newborns and those born after 37 completed gestational weeks. In addition, the point estimates have the wrong sign for all of our health measures. Since these coefficients represent an intention-to-treat effect of the week-37 rule around the cutoff, our estimates suggest that this rule yields no expected health benefits for newborns with gestational age close to 37 weeks.

The last row of each Panel in Table 2 presents the IV estimates of our structural equation. Not surprisingly, the estimated IV coefficients also point to a lack of short term health benefits for births supervised by obstetricians. While some of our coefficients have fairly large confidence intervals, as we detail in section 5.4, the point estimates remain consistently positive and generally insignificant in all our specification checks. The statistical insignificance of the coefficients combined with their consistently "wrong" signs suggest that

OB/GYN supervision provides little short-term health benefits for low-risk births around 37 completed gestational weeks. Since our results are similar across different measures of OB/GYN supervision, in the remainder of the paper we present results using our preferred measure—OB/GYN supervision from the onset of labor and including referrals for prematurity during delivery. Similar results for the other two measures are provided in tables A3–A6 in the Appendix.

## 5.4 Robustness Checks

In this section we investigate the robustness of our results to several scenarios that could lead to biased estimates in our framework. We start by checking the sensitivity of our results to the estimating strategy. If the key assumption in our RD design is satisfied (i.e., the variation in obstetrician supervision is as good as random around the week-37 cutoff), then including additional covariates in our model should not change our conclusions. In panel A of Table 3 we present IV estimates from a specification that includes the full set of controls described in the data section. We again find statistically insignificant adverse effects of obstetrician-supervision on newborn health. While the magnitude of the estimated effects is somewhat smaller than our baseline results, we cannot reject that the two sets of estimates are statistically equivalent.

Next, we turn to the possibility that our results could be driven by heaping at the cutoff. In order to address this issue, Barreca et al. (2011) suggest estimating "donut" regressions that exclude the observations at the cutoff. Panel B of Table 3 shows IV results estimated on a sample excluding newborns with gestational ages of 258 and 259 days. This strategy does not alter the main conclusion from our results, confirming that our findings are not driven by heaping at the cutoff point.

Panel C tests the sensitivity of our results to the degree of the polynomial in gestational age. Our choice of a linear function in gestational age is motivated by the reduced form relationship between infant outcomes and gestational age plotted in Figure 4, which does not indicate nonlinearities within

our bandwidth. When we reestimate our baseline regressions using a second degree polynomial (as before, allowed to vary on either side of the cutoff), the coefficients again indicate that births supervised by an obstetrician do not exhibit lower mortality.

In Panel D we test the robustness of our results to different bandwidths using intervals of 7 and 21 days on either side of the cutoff. The estimated effect of obstetrician supervision is insignificant and statistically indistinguishable from the baseline results regardless of which bandwidth we use.

We next turn to the choice of kernel. In Panel E of Table 3 we report results based on a triangular kernel which places less weight on observations farther away from the cutoff. Our results again point to no effects of OB/GYN supervision on infant health.

Mortality and low Apgar scores are rare events and estimating linear probability models for these outcomes could be problematic. For that reason, in Panel F we present results from non-linear models: we estimate the first stage and the reduced form regressions by probit, separately on each side of the cutoff; we calculate the average predicted values at the cutoff when approaching it from the left and from the right; and we calculate the Wald estimate as the ratio of the difference in average predicted outcomes at the cutoff to the difference in average predicted probability of obstetrician supervision at the cutoff (Hahn et al., 2001). The standard errors are obtained via bootstrap with 500 replications. The results are very similar to our baseline estimates, confirming that our main conclusions are not driven by nonlinearities.

In the remainder of the section we check the robustness of our findings to sample selection. Panel A of Table 4 examines the sensitivity of our results to different definitions of the running variable. The week-37 rule requires pregnant women to be referred to an obstetrician if labor occurs before 37 completed gestational weeks. Since we only observe the precise moment of the end of labor (i.e. the birth hour of the child), but not the exact moment of onset of labor, our baseline strategy defines the running variable as gestational age at birth measured in full days. We now define the running variable as gestational age at the onset of labor assuming various intervals between the

onset of labor and the actual birth.[26] For example, an infant born at 9am on gestational day 259 is not classified as premature according to our original definition of the running variable or when assuming a 6-hour labor, but it is classified as premature when considering a 12-hour labor. The results using these new definitions are virtually identical to our baseline estimates.

Recall that our analytic sample includes women either under the supervision of a midwife at the onset of labor or coded as referred to an obstetrician during pregnancy but at most one day before the birth. In Panel B of Table 4, we check whether the baseline results are sensitive to the subset of women who were referred during pregnancy included in the sample. Our analysis rests on the assumption that the prenatal care of women is the same, regardless of the type of medical attendant supervising the birth, so that the only difference is in the medical attendant. A shorter window between referral and delivery increases the likelihood of similar prenatal care. Therefore, row 1 restricts the subset of referrals during pregnancy to those on the same day as the delivery (i.e., excluding women referred one day earlier). The results are virtually the same as our baseline. In contrast, in the second row we add to our analysis sample all women referred to an OB/GYN for imminent prematurity, regardless of the date of referral (some women may be referred for prematurity several days before the actual delivery). Our results are again robust.

As mentioned in section 3, the date of referral is missing for a number of observations. So far we assumed that data are missing at random and thus ignored these observations. However, our results would be biased if the pattern of missing data is correlated with infant health and with prematurity. In order to shed light on this issue, in Panel C we check the sensitivity of our results to the inclusion of these observations. Since we do not observe the referral date, we are uncertain about the attending medical professional at the onset of labor. Therefore, we classify these observations as attended by the recorded medical professional (row 1), by a midwife (row 2) or by an OB/GYN (row 3). In all cases, the results are larger and indicate no health benefits from

[26]Note that we still use a 14-day bandwidth around the cutoff but the estimating sample changes because the cutoff changes with the definition of the running variable.

obstetrician supervision.

Finally, in Table 5 we focus on deliveries where the gain from OB/GYN supervision is presumably higher because of maternal and infant characteristics generally associated with worse outcomes. We first note that the first stage in all cases is as strong as in the overall analysis sample. Next, we find no health benefits from OB/GYN supervision among children with low birth weight, mothers above the age of 30, non-Dutch mothers and first births.[27] We do find that OB/GYN supervision is associated with better health outcomes among women living in a postal code in the first quartile of the distribution of average household income. However, this effect is not found in the second quartile, where the point estimate has the same sign as our baseline results. Moreover, previous research showed that home births (which are supervised by midwives and are not allowed before week 37) lead to increased infant mortality among individuals living in low-income neighborhoods (Daysal et al., 2012). Therefore, it is likely that this particular estimate is due to home births leading to worse outcomes among births supervised by midwives and not to a health benefit from obstetrician supervision.

In conclusion, we do not find any evidence that our results are driven by the specification of our RD design, by our sample selection, or by the construction of our running variable.

## 5.5 Compliers

The estimated coefficients in an instrumental variable framework represent a local average treatment effect (LATE) that applies to compliers: individuals who receive the treatment only because of the instrument. In our case, the compliers are births that are supervised by an OB/GYN only because of the week-37 rule, i.e. because of prematurity, but that would stay under the supervision of a midwife if the onset of labor was after 258 gestational days. This LATE is interesting in itself because of several reasons. First, it evaluates

[27] We might expect stronger health benefits for first births because fewer potential risk factors might be known. For instance, the List of Obstetric Indications defines a number of risk factors based on previous pregnancies.

the benefits of a relevant medical policy. Second, compliers comprise a large share of the sample. While we cannot identify individual compliers, we can calculate their share in the analysis sample as the first stage coefficient on the instrument (Angrist and Pischke, 2009). Based on the estimates in Table 2, repeated in the last row of Table 6, we find that compliers represent almost 60 percent of our analysis sample. Finally, compliers have characteristics similar to the overall sample. Table 6 shows the likelihood that a birth in a particular subgroup is a complier relative to the overall sample.[28] This is calculated as the ratio of the first stage coefficient in that particular subgroup to the first stage coefficient in the analysis sample. Our findings indicate that compliers are similar to the average women in the sample based on an array of characteristics of the mother (ethnicity, age, intended place of birth), of the newborn (parity, birth weight) and of the mother's postal code of residence (average household income) because all the likelihood ratios are very close to 1.[29]

## 5.6 Medical Treatments

Recall that our estimates combine the effects of OB/GYN supervision and of the additional treatments that medical doctors can provide. While we cannot reliably identify in our data all the treatments that newborns receive, we do have a reliable measure of one specific treatment: admission to a neonatal intensive care unit (NICU). Table 7 shows the estimates of our RD strategy applied to NICU admissions. Admission rates to NICUs are high around the cutoff: about 8 percent to the right and around 34 percent to the left. Our reduced form estimates indicate that the week-37 rule induces an increase in NICU admission of about 15 percentage points and our IV-results show that infants delivered under the supervision of an OB/GYN are 25 percentage points more likely to be admitted to a NICU. This effect is robust across the distribution of average household income, consistent with our earlier inter-

[28] Alternatively, the average characteristics of compliers can be calculated using the methodology in Almond and Doyle (2011). Both methods provide the same qualitative results.

[29] The share of compliers changes depending on the definition of the treatment but their characteristics remain similar to the overall sample (see Appendix Table A10).

pretation that OB/GYN supervision is not particularly different among those living in poorer areas. The fact that NICU admission rates are significantly higher among OB/GYN supervised deliveries indicates the potential for significant cost savings by shifting low-risk deliveries from obstetricians to midwives. These savings come not only from the use of lower-cost physician extenders but also from the elimination of potentially excessive utilization of medical treatments provided by OB/GYNs.

# 6 Conclusions

In this paper, we examine the impact of OB/GYN (as opposed to midwife) supervision on the health outcomes of low-risk newborns. In order to address the endogeneity in obstetrician supervision, we exploit the exogenous variation in the medical professional attending the delivery generated by a policy rule in the Netherlands. The policy rule requires that low-risk women give birth under the supervision of a midwife unless the birth occurs before 37 completed gestational weeks. This motivates the use of a regression discontinuity design to estimate the causal impact of OB/GYN supervision on newborn outcomes.

Using data from the Netherlands for the period 2000–2008, we find that the policy rule leads to a statistically and economically large increase in the probability of obstetrician-supervision below the week-37 cutoff. We empirically confirm the validity of our RD design by showing that there are no discontinuities at the week-37 cutoff in the frequency of births or in a wide range of observable characteristics pertaining to mothers and newborns. Despite the substantial variation in OB/GYN supervision, our results indicate that average newborn health outcomes are remarkably similar across the week-37 cutoff. In addition, we find that obstetrician-supervised deliveries lead to considerable higher rates of NICU admission. Therefore, midwife supervision of low-risk deliveries can lead to substantial cost reductions without compromising newborn health through reduced use of medical technologies combined with lower personnel costs.

A few limitations to our study should be noted. First, our results apply to

the sample of compliers, i.e., individuals who are supervised by an OB/GYN instead of a midwife only because of a premature birth. However, compliers represent a large share of the sample and they seem to be representative for the entire sample. In addition, these deliveries are slightly riskier than at-term deliveries based on their shorter gestational age. Hence, it is likely that our results on the safety of midwife supervision generalize to at-term deliveries with higher gestational ages. Second, we only focus on newborn short-term outcomes because of data availability. Although we do not find evidence of benefits from OB/GYN supervision in terms of these outcomes, it is possible that there are long-term benefits to newborns or benefits to maternal health that we cannot capture with our measures. Finally, the study uses data from the Netherlands, whose obstetric care is based on a well-established system of risk classification and referrals. This system ensures that high-risk deliveries are identified and referred to obstetricians for supervision, but also that midwives receive extensive training preparing them to supervise deliveries on their own, without any OB/GYN present. Therefore, our findings should be interpreted with caution in settings where risk assessment is less developed or where midwives receive less training.

Given the steep increase in the health care costs for the very young (i.e., infants under 1 year old), the possibility of reducing costs through the use of physician extenders such as midwifery care is likely to be an important policy topic in the coming years. Taken together, our results point to potential cost savings from increased use of midwifery care for low-risk deliveries.

# References


Almond, Douglas, and Joseph Doyle (2011), "After midnight: A regression discontinuity design in length of postpartum hospital stays," *American Economic Journal: Economic Policy* 3(3), 1–34.

Almond, Douglas, Joseph Doyle, Amanda Kowalski, and Heidi Williams (2010), "Estimating marginal returns to medical care: Evidence from at-risk newborns," *Quarterly Journal of Economics* 125(2), 591–634.

——— (2011), "The role of hospital heterogeneity in measuring marginal returns to medical care: A reply to Barreca, Guldi, Lindo, and Waddell," *The Quarterly Journal of Economics* .

Amelink-Verburg, Marianne, and Simone Buitendijk (2010), "Pregnancy and labour in the Dutch maternity care system: What is normal? The role division between midwives and obstetricians," *Journal of Midwifery & Women's Health* 55(3), 216–225.

Angrist, Joshua, and Jorn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell (2011), "Heaping-induced bias in regression-discontinuity designs," NBER Working Paper no. 17408.

Bharadwaj, Prashant, Katrine Løken, and Christopher Neilson (forthcoming), "Early life health interventions and academic achievement," *American Economic Review* .

Birthplace in England Collaborative Group (2011), "Perinatal and maternal outcomes by planned place of birth for healthy women with low risk pregnancies: the Birthplace in England national prospective cohort study," *BMJ* 343, d7400.

Chambliss, L.R., C. Daly, A.L. Medearis, M. Ames, M. Kayne, and R. Paul (1992), "The role of selection bias in comparing cesarean birth rates between physician and midwifery management," *Obstetrics & Gynecology* 80(2), 161–165.

Chandra, Amitabh, and Jonathan Skinner (2012), "Technology growth and expenditure growth in health care," *Journal of Economic Literature* 50(3), 645–680.

Currie, Janet, and W. Bentley MacLeod (2008), "First do no harm? Tort reform and birth outcomes," *Quarterly Journal of Economics* 123(2), 795–830.

Cutler, David, Mark McClellan, Joseph Newhouse, and Dahlia Remler (1998), "Are medical prices declining? Evidence from heart attack treatments," *Quarterly Journal of Economics* 113(4), 991–1024.

Cutler, David, and Ellen Meara (1998), "The medical costs of the young and old: A forty-year perspective," in David Wise, ed., *Frontiers in the Economics of Aging*, 215–246, University of Chicago Press.

——— (2000), "The technology of birth: Is it worth it?" *NBER/Frontiers in Health Policy Research* 3(1), 33–67.

CVZ (2003), *Verloskundig Vademecum*, Diemen, The Netherlands.

Daysal, N. Meltem, Mircea Trandafir, and Reyn van Ewijk (2012), "Saving lives at birth: The impact of home births on infant outcomes," IZA Discussion Paper no. 6879.

de Jonge, A., B.W. Mol, B.Y. van der Goes, J.G. Nijhuis, J.A. van der Post, and S.E. Buitendijk (2010), "Too early to question effectiveness of dutch system," *BMJ* 341.

de Jonge, A., B.Y. van der Goes, A.C.J. Ravelli, M.P. Amelink-Verburg, B.W. Mol, J.G. Nijhuis, J. Bennebroek Gravenhorst, and S.E. Buitendijk (2009), "Perinatal mortality and morbidity in a nationwide cohort of 529,688 low-risk planned home and hospital births," *BJOG: An International Journal Of Obstetrics And Gynaecology* 116(9), 1177–84.

Evers, Annemieke, Hens Brouwers, Chantal Hukkelhoven, Peter Nikkels, Janine Boon, Anneke van Egmond-Linden, Jacqueline Hillegersberg, Yvette Snuif, Sietske Sterken-Hooisma, Hein Bruinse, and Anneke Kwee (2010), "Perinatal mortality and severe morbidity in low and high risk term pregnancies in the Netherlands: Prospective cohort study," *BMJ (Clinical research ed.)* 341, c5639.

Freedman, Seth (2012), "The effect of deregionalization on health outcomes: Evidence from neonatal intensive care," Mimeo.

Fuchs, Victor (1998), *Who Shall Live?: Health, Economics, and Social Choice*, World Scientific.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw (2001), "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica* 69(1), 201–209.

Harvey, S., J. Jarrell, R. Brant, C. Stainton, and D. Rach (2007), "A randomized, controlled trial of nurse-midwifery care," *Birth* 23(3), 128–135.

Imbens, Guido, and Thomas Lemieux (2008), "Regression discontinuity designs: A guide to practice," *Journal of Econometrics* 142(2), 615–635.

Institute of Medicine (2007), *Preterm Birth: Causes, Consequences, and Prevention*, Washington, D.C.: The National Academies Press.

——— (2011), *The Future of Nursing: Leading Change, Advancing Health*, Washington, D.C.: The National Academies Press.

Janssen, P.A., S.K. Lee, E.M. Ryan, D.J. Etches, D.F. Farquharson, D. Peacock, and M.C. Klein (2002), "Outcomes of planned home births versus planned hospital births after regulation of midwifery in British Columbia," *Canadian Medical Association Journal* 166(3), 315–323.

Kochanek, K.D., S.E. Kirmeyer, J.A. Martin, D.M. Strobino, and B. Guyer (2012), "Annual summary of vital statistics: 2009," *Pediatrics* 129(2), 338–348.

Kramer, Michael, Aris Papageorghiou, Jennifer Culhane, Zulfiqar Bhutta, Robert Goldenberg, Michael Gravett, Jay Iams, Agustin Conde-Agudelo, Sarah Waller, Fernando Barros, Hannah Knight, and Jose Villar (2012), "Challenges in defining and classifying the preterm birth syndrome," *American Journal of Obstetrics and Gynecology* 206(2), 108–112.

Lee, David, and David Card (2008), "Regression discontinuity inference with specification error," *Journal of Econometrics* 142(2), 655–674.

Lee, David, and Thomas Lemieux (2010), "Regression discontinuity designs in economics," *Journal of Economic Literature* 48(2), 281–355.

MacDorman, M.F., and G.K. Singh (1998), "Midwifery care, social and medical risk factors, and birth outcomes in the USA," *Journal of Epidemiology and Community Health* 52(5), 310–317.

McClellan, Mark, and Joseph Newhouse (1997), "The marginal cost-effectiveness of medical technology: A panel instrumental-variables approach," *Journal of Econometrics* 77(1), 39–64.

McClellan, Mark, and Haruko Noguchi (1998), "Technological change in heart-disease treatment. Does high tech mean low value?" *American Economic Review* 88(2), 90–96.

McCrary, Justin (2008), "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics* 142(2), 698–714.

Miller, Amalia (2006), "The impact of midwifery-promoting public policies on medical interventions and health outcomes," *B.E. Journal of Economic Analysis and Policy: Advances in Economic Analysis and Policy* 6(1), 1–34.

Nationale Atlas Volksgezondheid (2011), "Reistijd naar ziekenhuis met afdeling verloskunde 2011," available online at http://www.zorgatlas.nl/zorg/ziekenhuiszorg/algemene-en-academische-ziekenhuizen/aanbod/reistijd-naar-ziekenhuis-met-afdeling-verloskunde, accessed Sept. 4, 2012.

Newhouse, Joseph (1992), "Medical care costs: How much welfare loss?" *Journal of Economic Perspectives* 6(3), 3–21.

OECD (2012), "Health data 2012. Health expenditure and financing," data extracted on Sept. 4, 2012.

Redshaw, M., R. Rowe, L. Schroeder, D. Puddicombe, A. Macfarlane, M. Newburn, C. McCourt, J. Sandall, L. Silverton, and N. Marlow (2011), "Mapping maternity care: the configuration of maternity care in England. Birthplace in England research programme. Final report part 3," NIHR Service Delivery and Organisation programme.

Rosenblatt, R.A., S.A. Dobie, L.G. Hart, R. Schneeweiss, D. Gould, T.R. Raine, T.J. Benedetti, M.J. Pirani, and E.B. Perrin (1997), "Interspecialty differences in the obstetric care of low-risk women," *American Journal of Public Health* 87(3), 344–351.

Sandall, J., H. Soltani, S. Gates, A. Shennan, and D. Devane (2013), "Midwife-led continuity models versus other models of care for childbearing women (review)," *The Cochrane Library* 8.

Schakel, W, and J Bekhof (2010), "Prematuren geboren na 36 weken zwangerschapsduur - 48 uur observatie op de kraamafdeling is voldoende [prematures born after 36 weeks of gestation – 48 hours of observation in the obstetric ward is sufficient]," *Tijdschrift voor kindergeneeskunde* 78, 3–6.

Skinner, Jonathan, Douglas Staiger, and Elliott Fisher (2006), "Is technological change in medicine always worth it? The case of acute myocardial infarction," *Health Affairs* 25(2), w34–w47.

Tumbull, D., A. Holmes, N. Shields, H. Cheyne, S. Twaddle, W.H. Gilmour, M. McGinley, M. Reid, I. Johnstone, I. Geer, et al. (1996), "Randomised, controlled trial of efficacy of midwife-managed care," *The Lancet* 348(9022), 213–218.

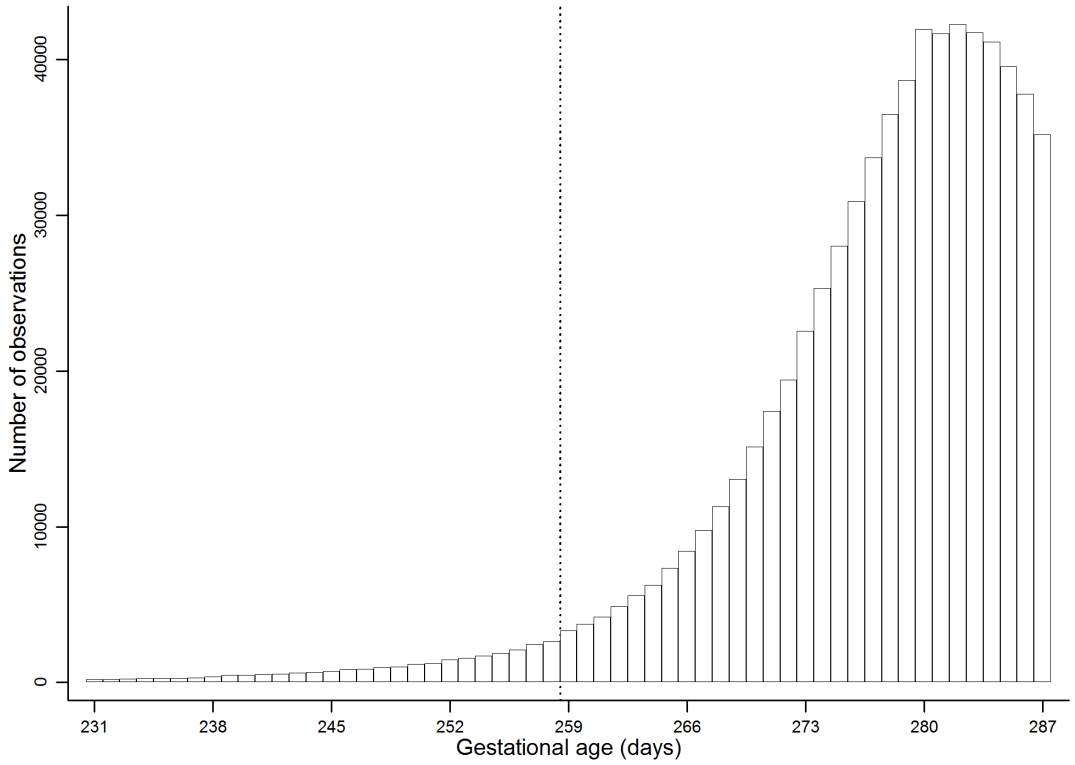
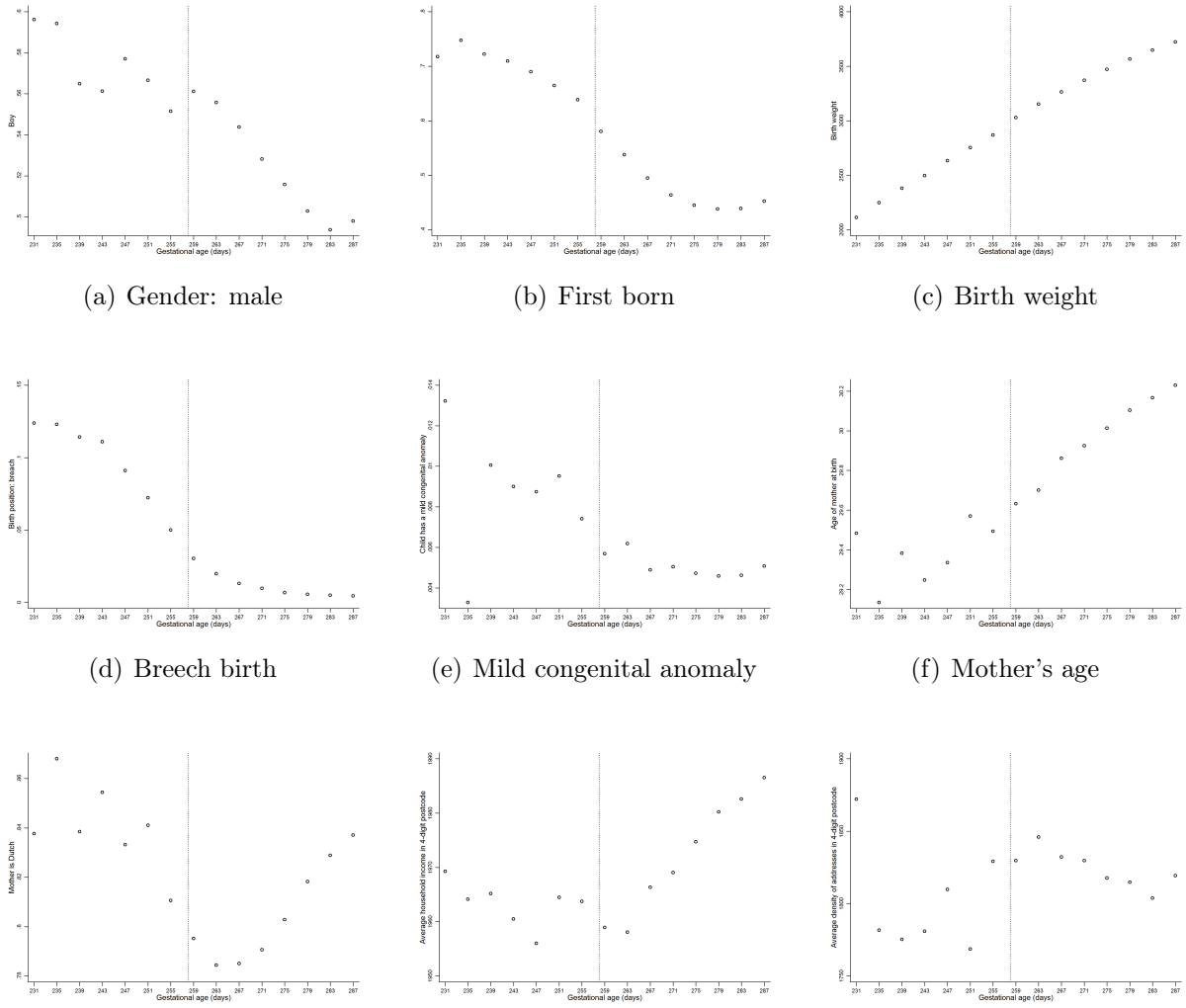



Figure 1: Frequency of births around 37 completed gestational weeks

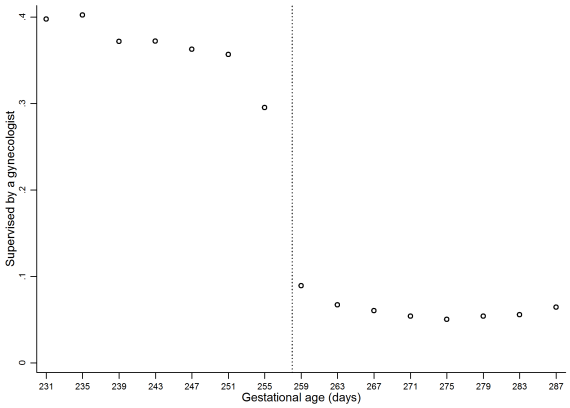



(a) Gender: male (b) First born (c) Birth weight

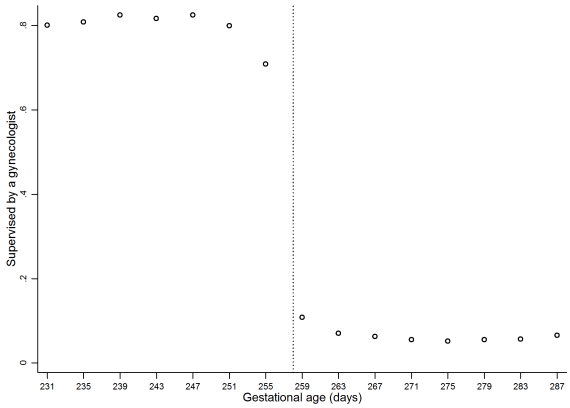
(d) Breech birth (e) Mild congenital anomaly (f) Mother's age

(g) Mother's ethnicity: Dutch (h) Average household income (i) Average density

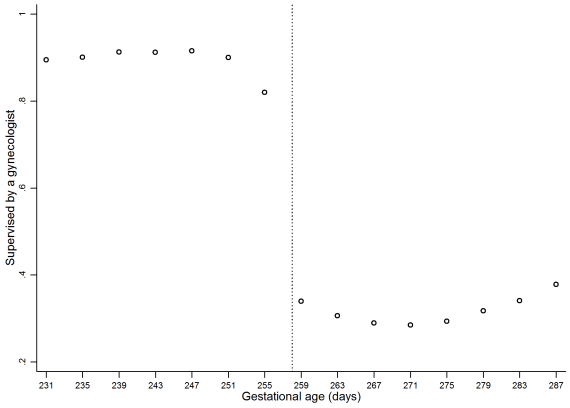
Figure 2: Evolution of selected covariates around the discontinuity




(a) From the onset of labor

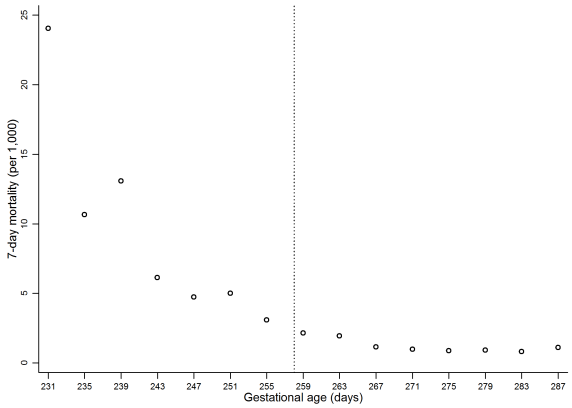



(b) From the onset of labor + referrals for prematurity during delivery

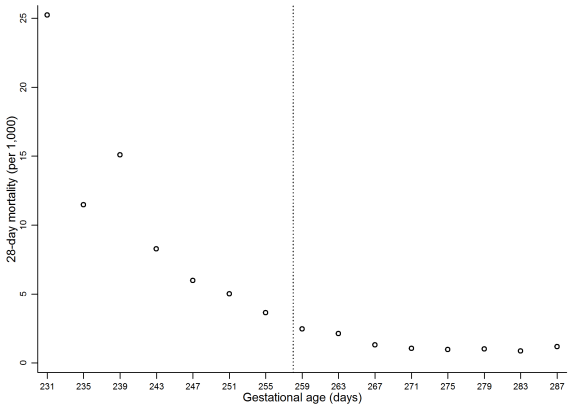



(c) At any point during delivery

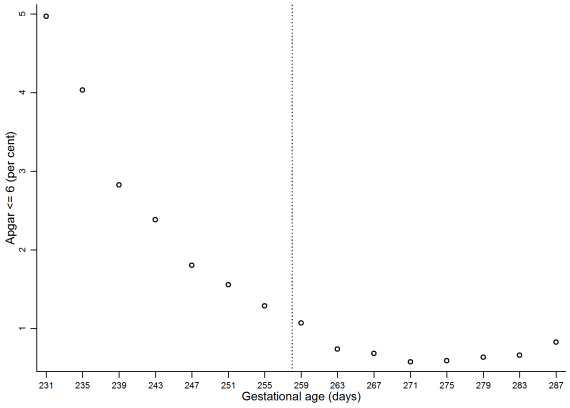
Figure 3: OB/GYN Supervision




(a) 7-day mortality




(b) 28-day mortality




(c) Low Apgar score

Figure 4: Newborn health

Table 1: Comparison of selected characteristics around the discontinuity

| | Gestational age | | Clustered |
|---|---|---|---|
| | Less than 37 completed weeks (1) | More than 37 completed weeks (2) | p-value for differences (3) |
| **A. OB/GYN supervision** | | | |
| From the onset of labor | 0.291 | 0.093 | 0.000 |
| From the onset of labor + referrals for prematurity during delivery | 0.696 | 0.110 | 0.000 |
| At any point during delivery | 0.809 | 0.346 | 0.000 |
| **B. Health outcomes** | | | |
| 7-day mortality (per 1,000) | 3.262 | 2.542 | 0.208 |
| 28-day mortality (per 1,000) | 3.377 | 2.837 | 0.402 |
| Low Apgar score (per cent) | 1.119 | 1.049 | 0.589 |
| **C. Maternal characteristics** | | | |
| Age | 29.565 | 29.536 | 0.625 |
| Ethnicity | | | |
| Dutch | 0.809 | 0.793 | 0.018 |
| Mediterranean | 0.076 | 0.075 | 0.832 |
| **D. Infant characteristics** | | | |
| Boy | 0.550 | 0.572 | 0.000 |
| Birth weight | 2,910 | 2,958 | 0.000 |
| Very low birth weight ($<$ 1,500g) | 0.0002 | 0.0001 | 0.624 |
| Low birth weight (1,500–2,500g) | 0.104 | 0.091 | 0.192 |
| First born | 0.629 | 0.605 | 0.000 |
| Congenital anomalies | | | |
| Mild | 0.008 | 0.007 | 0.210 |
| Severe | 0.014 | 0.011 | 0.199 |
| Multiple birth | 0.001 | 0.002 | 0.102 |
| Birth position | | | |
| Breech birth | 0.043 | 0.033 | 0.001 |
| Other | 0.025 | 0.028 | 0.210 |
| **E. Postal code characteristics** | | | |
| Average household income (euros) | 1,964 | 1,954 | 0.027 |
| Average density | 1,824 | 1,840 | 0.491 |
| Average percent 0–15 year-old | 18.985 | 19.055 | 0.202 |
| Number of observations | 20,566 | 129,905 | |

Notes: Each cell represents the mean of the corresponding variable in the row after controlling for gestational age. The last column presents the p-value for differences in means clustered at the gestational day level.

Table 2: Baseline results

| | 7-day mortality (per 1,000) (1) | 28-day mortality (per 1,000) (2) | Low Apgar score (per cent) (3) |
|---|---|---|---|
| **A. OB/GYN supervision from the onset of labor** | | | |
| First stage: $W37$ | 0.198*** | 0.198*** | 0.199*** |
| | (0.017) | (0.017) | (0.017) |
| Reduced form: $W37$ | 0.720 | 0.540 | 0.070 |
| | (0.572) | (0.645) | (0.130) |
| Instrumental variable: $OB/GYN$ | 3.631 | 2.725 | 0.352 |
| | (2.940) | (3.368) | (0.644) |
| **B. OB/GYN supervision from the onset of labor + referrals for prematurity** | | | |
| First stage: $W37$ | 0.586*** | 0.586*** | 0.587*** |
| | (0.042) | (0.042) | (0.042) |
| Reduced form: $W37$ | 0.720 | 0.540 | 0.070 |
| | (0.572) | (0.645) | (0.130) |
| Instrumental variable: $OB/GYN$ | 1.228 | 0.922 | 0.119 |
| | (1.012) | (1.134) | (0.218) |
| **C. OB/GYN supervision at any point during delivery** | | | |
| First stage: $W37$ | 0.463*** | 0.463*** | 0.464*** |
| | (0.036) | (0.036) | (0.036) |
| Reduced form: $W37$ | 0.720 | 0.540 | 0.070 |
| | (0.572) | (0.645) | (0.130) |
| Instrumental variable: $OB/GYN$ | 1.555 | 1.167 | 0.151 |
| | (1.291) | (1.438) | (0.275) |
| Average of health outcome after gestational day 258 | 1.447 | 1.617 | 0.710 |
| Observations | 150,471 | 150,471 | 150,269 |

Notes: Each cell represents a different regression. All specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS. Sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Robust standard errors clustered at the gestational day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Robustness to estimation strategy

| | 7-day mortality (per 1,000) (1) | 28-day mortality (per 1,000) (2) | Low Apgar score (per cent) (3) |
|---|---|---|---|
| **A. Including controls** | 0.851 | 0.495 | 0.047 |
| | (0.999) | (1.086) | (0.226) |
| Observations | 150,471 | 150,471 | 150,269 |
| **B. Donut sample (excluding days 258 and 259)** | 0.652 | 0.180 | 0.218 |
| | (1.093) | (1.269) | (0.245) |
| Observations | 144,520 | 144,520 | 144,327 |
| **C. Sensitivity to polynomial in gestational age** | | | |
| **Second degree polynomial** | 0.886 | 1.972 | −0.116 |
| | (1.463) | (1.335) | (0.362) |
| Observations | 150,471 | 150,471 | 150,269 |
| **Third degree polynomial** | 2.867 | 4.552** | −0.895* |
| | (2.301) | (1.795) | (0.480) |
| Observations | 150,471 | 150,471 | 150,269 |
| **D. Sensitivity to bandwidth choice** | | | |
| **7-day bandwidth** | 1.337 | 2.113 | −0.229 |
| | (1.467) | (1.322) | (0.342) |
| Observations | 49,163 | 49,163 | 49,088 |
| **21-day bandwidth** | 0.556 | 0.419 | 0.289 |
| | (1.006) | (1.170) | (0.224) |
| Observations | 369,740 | 369,740 | 369,266 |
| **E. Triangular kernel** | −0.261 | 0.020 | 0.372 |
| | (1.289) | (1.472) | (0.304) |
| Observations | 150,471 | 150,471 | 150,269 |
| **F. Non-linear specification** | | | |
| **First stage:** $W37$ | 0.578*** | 0.578*** | 0.579*** |
| | (0.006) | (0.006) | (0.006) |
| **Reduced form:** $W37$ | 0.562 | 0.458 | 0.068 |
| | (0.776) | (0.820) | (0.148) |
| **IV:** $OB/GYN$ | 0.972 | 0.793 | 0.118 |
| | (1.342) | (1.418) | (0.257) |
| Observations | 150,471 | 150,471 | 150,269 |

Notes: Each cell represents a different regression. Unless otherwise indicated, all specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS; sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Robust standard errors clustered at the gestational day level, except for Panel F which reports bootstrapped standard errors from 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Robustness to sample definition

| | 7-day mortality (per 1,000) (1) | 28-day mortality (per 1,000) (2) | Low Apgar score (per cent) (3) |
|---|---|---|---|
| **A. Alternative definitions of the running variable** | | | |
| **Gestational age at birth** | 1.414 | 1.211 | 0.110 |
| **minus 6 hours** | (1.134) | (1.102) | (0.309) |
| Observations | 155,916 | 155,916 | 155,707 |
| **Gestational age at birth** | 1.468 | 1.205 | 0.172 |
| **minus 12 hours** | (1.244) | (1.255) | (0.310) |
| Observations | 161,978 | 161,978 | 161,764 |
| **B. Inclusion of referrals to OB/GYN prior to the onset of labor** | | | |
| **Same day as delivery** | 1.257 | 0.640 | 0.211 |
| | (1.113) | (1.282) | (0.242) |
| Observations | 145,549 | 145,549 | 145,351 |
| | (1.046) | (1.230) | (0.240) |
| Observations | 147,301 | 147,301 | 147,101 |
| **All referrals for prematurity** | 0.755 | 0.450 | 0.049 |
| | (0.975) | (1.119) | (0.214) |
| Observations | 151,351 | 151,351 | 151,148 |
| **C. Including observations with missing referral dates** | | | |
| **Supervised by recorded attendant** | 1.590 | 2.460* | 0.160 |
| | (1.167) | (1.285) | (0.311) |
| **All supervised by midwife** | 1.909 | 2.954* | 0.192 |
| | (1.421) | (1.563) | (0.374) |
| **All supervised by OB/GYN** | 5.858 | 9.064* | 0.590 |
| | (4.128) | (4.550) | (1.132) |
| Observations | 226,815 | 226,815 | 226,514 |

Notes: Each cell represents a different regression. Unless otherwise indicated, all specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS; sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Robust standard errors clustered at the gestational day level, except for Panel F which reports bootstrapped standard errors from 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Heterogeneous effects

| | First stage | 7-day mortality (per 1,000) | 28-day mortality (per 1,000) | Low Apgar score (per cent) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Baseline results** | 0.586*** | 1.228 | 0.922 | 0.119 |
| | (0.042) | (1.012) | (1.134) | (0.218) |
| Observations | | 150,471 | 150,471 | 150,269 |
| Average health outcome | | 1.447 | 1.617 | 0.710 |
| **A. Mother: Non-Dutch** | 0.569*** | −2.775 | −2.275 | −0.171 |
| | (0.033) | (3.221) | (4.117) | (0.375) |
| Average health outcome | | 1.694 | 1.911 | 0.888 |
| Observations | | 31,293 | 31,293 | 31,259 |
| **B. Mother: Older than median (30 years)** | 0.559*** | 1.425 | 0.500 | 0.155 |
| | (0.044) | (2.025) | (1.875) | (0.508) |
| Average health outcome | | 1.525 | 1.660 | 0.703 |
| Observations | | 68,372 | 68,372 | 68,278 |
| **C. Infant: Low birth weight (1,500-2,500g)** | 0.534*** | 3.544 | 3.209 | 0.656 |
| | (0.035) | (6.024) | (6.003) | (1.113) |
| Observations | | 9,440 | 9,440 | 9,424 |
| Average health outcome | | 6.917 | 7.350 | 2.383 |
| **D. Infant: First birth** | 0.614*** | 1.085 | 1.104 | 0.043 |
| | (0.036) | (0.774) | (0.987) | (0.261) |
| Observations | | 79,541 | 79,541 | 79,431 |
| Average health outcome | | 1.425 | 1.623 | 0.865 |
| Average health outcome | | 1.470 | 1.610 | 0.549 |
| **E. Average household income** | | | | |
| **First quartile** | 0.565*** | −6.501*** | −6.527*** | −0.751* |
| | (0.044) | (1.724) | (2.284) | (0.416) |
| Observations | | 37,716 | 37,716 | 37,659 |
| Average health outcome | | 1.746 | 1.868 | 0.899 |
| **Second quartile** | 0.595*** | 3.739 | 2.742 | 0.404 |
| | (0.044) | (2.358) | (2.998) | (0.615) |
| Observations | | 37,536 | 37,536 | 37,476 |
| Average health outcome | | 1.145 | 1.300 | 0.648 |

Notes: Each cell represents a different regression. All specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS; sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Column 1 lists the first stage coefficient of $W37$ in the the mortality sample; columns 2-4 the coefficient of $OB/GYN$ in the structural equation. Robust standard errors clustered at the gestational day level. The average health outcome is for observations to the right of the cutoff (gestational age greater than 258 days). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Complier characteristics, relative likelihoods

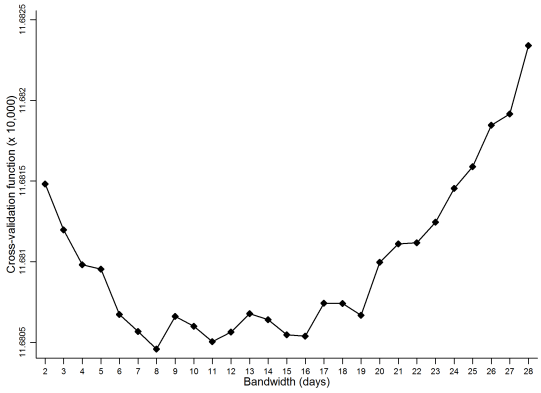
| | Relative likelihood (1) | Observations (2) |
|---|---|---|
| **A. Maternal characteristics** | | |
| Ethnicity: Dutch | 1.007 | 119,178 |
| Age | | |
| Less than 35 | 1.015 | 133,736 |
| Less than median (30 years) | 1.036 | 82,099 |
| Intended home birth | 0.994 | 55,981 |
| **B. Infant characteristics** | | |
| Birth weight | | |
| Less than median (3,160g) | 1.010 | 75,980 |
| Low birth weight (1,500–2,500g) | 0.910 | 9,440 |
| First born | 1.046 | 79,541 |
| **C. Postal code characteristics** | | |
| Average household income | | |
| Below median (1,921 euros) | 0.989 | 75,252 |
| First quartile | 0.964 | 37,716 |
| Second quartile | 1.014 | 37,536 |
| Third quartile | 1.049 | 37,633 |
| Fourth quartile | 0.973 | 37,586 |
| Share of compliers | 0.586 | 150,471 |

Notes: Column 1 shows the likelihood that compliers have the characteristic indicated in the row relative to the entire analysis sample within our chosen bandwidth. Column 2 shows the number of observations with that characteristic in the analysis sample. The last row gives the share of compliers (based on the first-stage in the mortality sample) and total number of observations in the analysis sample.

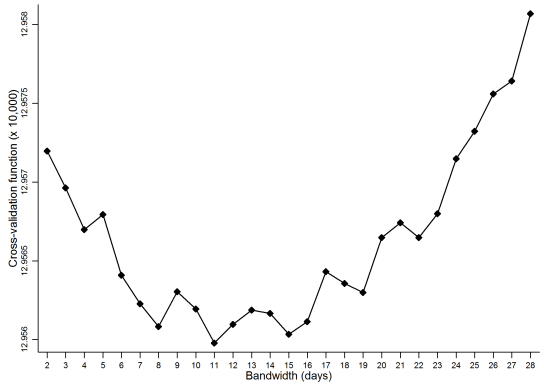
Table 7: NICU admissions (per cent)

| | All | Average Household Income | | | |
|---|---|---|---|---|---|
| | | 1st quartile | 2nd quartile | 3rd quartile | 4th quartile |
| | (1) | (2) | (3) | (4) | (5) |
| **First stage:** $W37$ | 0.586*** | 0.565*** | 0.595*** | 0.615*** | 0.571*** |
| | (0.042) | (0.044) | (0.044) | (0.040) | (0.043) |
| **Reduced form:** $W37$ | 14.774*** | 13.378*** | 14.424*** | 16.467*** | 14.857*** |
| | (1.198) | (1.974) | (1.169) | (0.819) | (1.869) |
| **Instrumental variable:** $OB/GYN$ | 25.194*** | 23.666*** | 24.252*** | 26.782*** | 26.034*** |
| | (0.857) | (2.046) | (1.008) | (1.303) | (2.878) |
| Average of outcome after gestational day 258 | 7.927 | 7.857 | 8.732 | 8.183 | 6.939 |
| Observations | 150,471 | 37,716 | 37,536 | 37,633 | 37,586 |

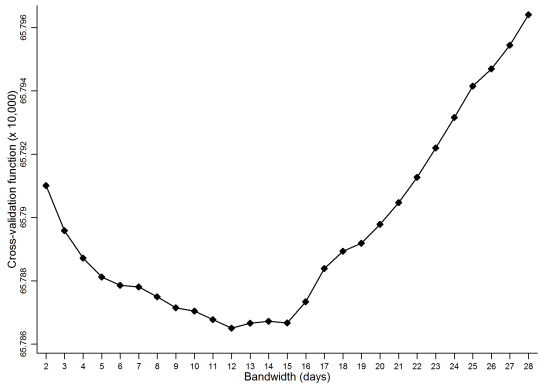
Notes: Each cell represents a different regression. All specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS. Sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Robust standard errors clustered at the gestational day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$




(a) 7-day mortality




(b) 28-day mortality




(c) Low Apgar score

Figure A1: Cross-validation function using observations on both sides of the discontinuity

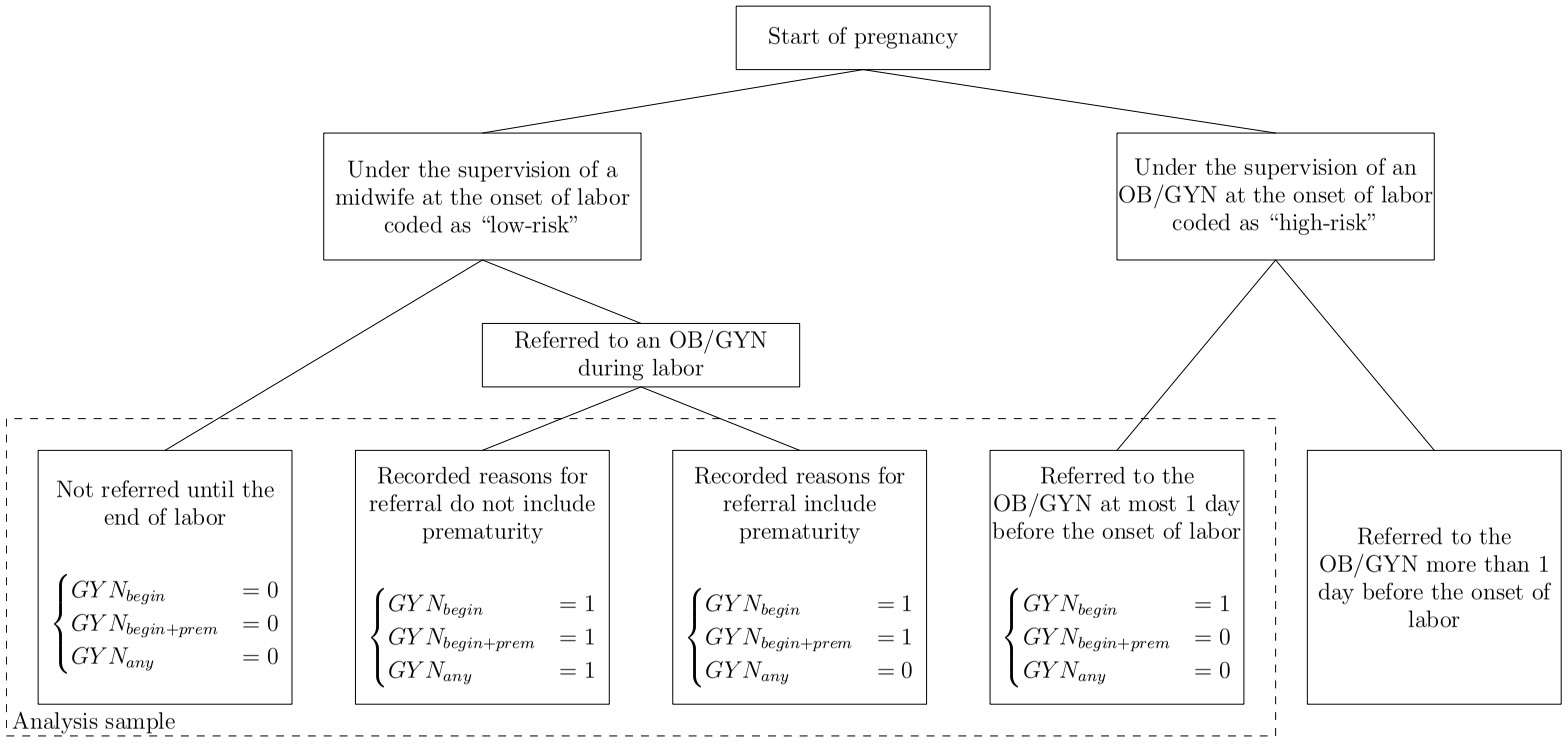
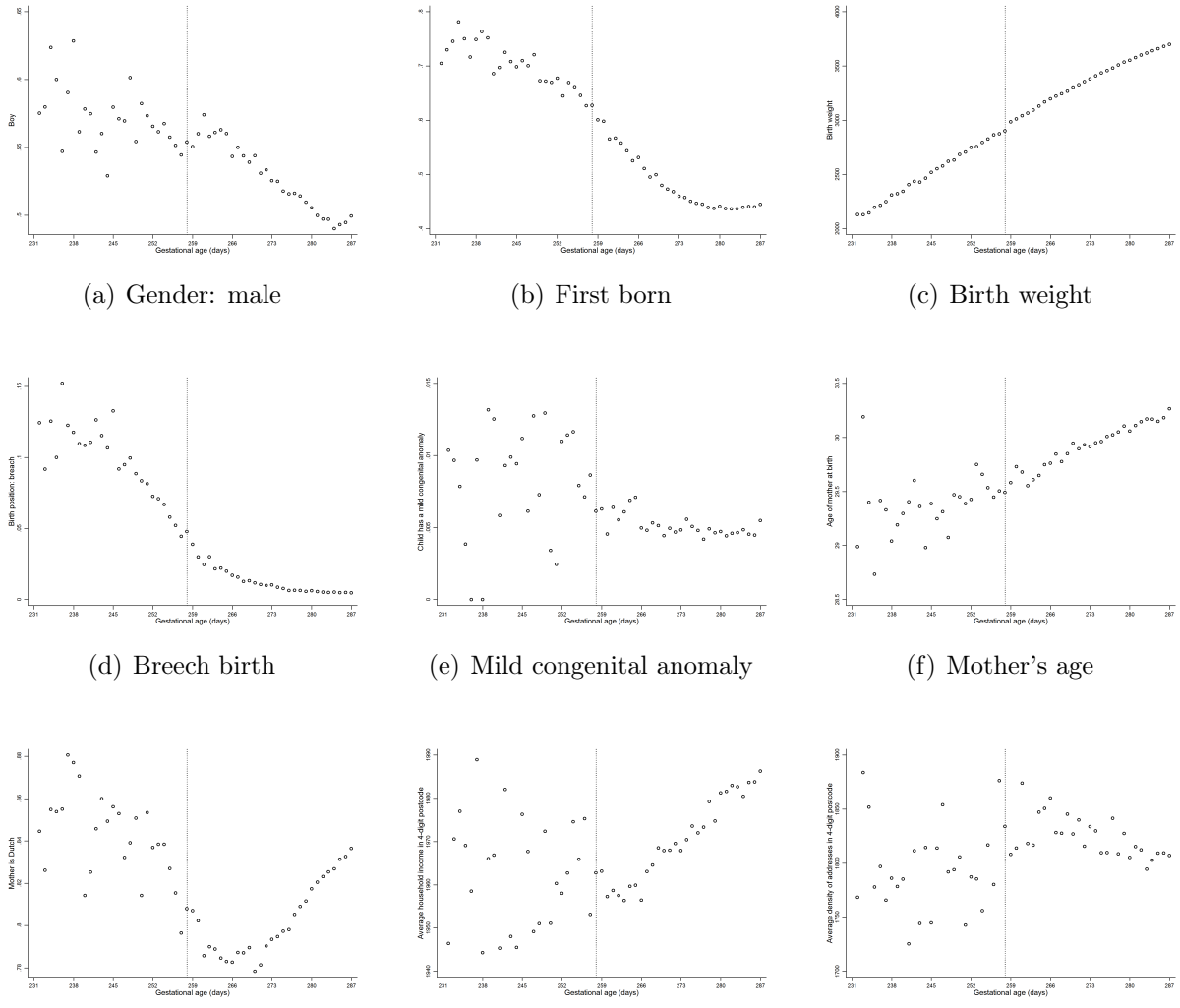



Figure A2: Definition of the three measures of OB/GYN supervision

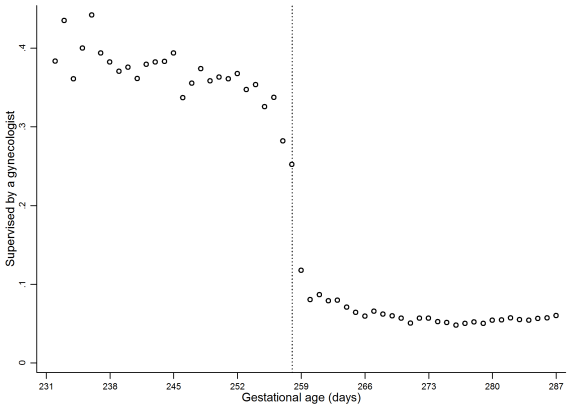



(a) Gender: male (b) First born (c) Birth weight

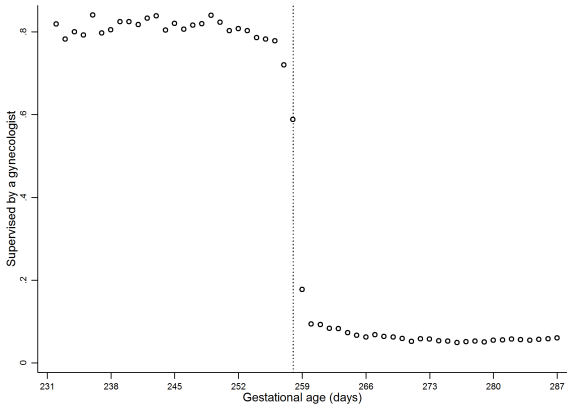
(d) Breech birth (e) Mild congenital anomaly (f) Mother's age

(g) Mother's ethnicity: Dutch (h) Average household income (i) Average density

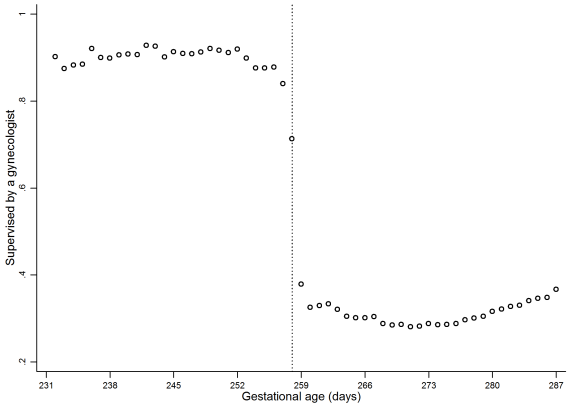
Figure A3: Evolution of selected covariates around the discontinuity




(a) From the onset of labor

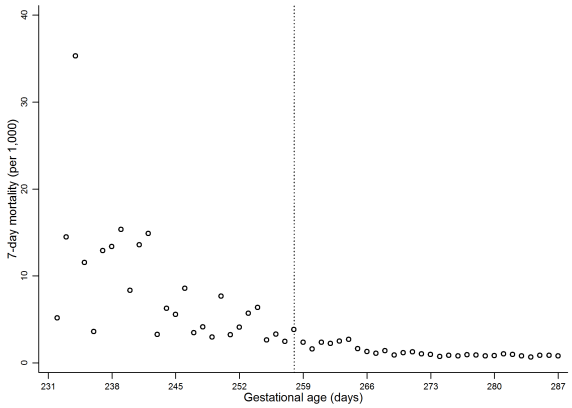



(b) From the onset of labor + referrals for prematurity during delivery

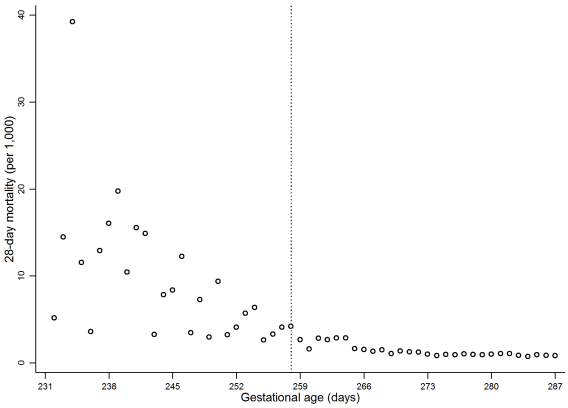



(c) At any point during delivery

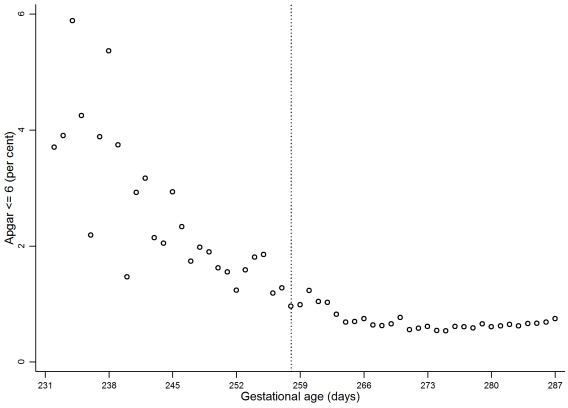
Figure A4: OB/GYN Supervision




(a) 7-day mortality




(b) 28-day mortality




(c) Low Apgar score

Figure A5: Newborn health

Table A1: Optimal bandwidth, gestational age in days

| | 7-day mortality (1) | 28-day mortality (2) | Low Apgar score (3) |
|---|---|---|---|
| **A. Rule-of-thumb bandwidth** | | | |
| Left | 9.33 | 8.88 | 12.08 |
| Right | 11.94 | 11.71 | 9.57 |
| Both | 6.68 | 6.34 | 9.52 |
| **B. Cross-validation bandwidth** | | | |
| Left | 8 | 11 | 15 |
| Right | 14 | 13 | 11 |
| Both | 8 | 11 | 12 |

Notes: See section 4.2 for details on the calculation of optimal bandwidths.

Table A2: Comparison of characteristics around the discontinuity

| | Gestational age | | Clustered |
|---|---|---|---|
| | Less than 37 completed weeks (1) | More than 37 completed weeks (2) | p-value for differences (3) |
| **A. Distribution of maternal age** | | | |
| 20–24 | 0.119 | 0.127 | 0.030 |
| 25–29 | 0.334 | 0.325 | 0.064 |
| 30–34 | 0.373 | 0.372 | 0.920 |
| 35–39 | 0.133 | 0.135 | 0.554 |
| 40 and above | 0.016 | 0.015 | 0.521 |
| **B. Variables with missing values** | | | |
| Mother's age | 0.0002 | 0.0001 | 0.639 |
| Newborn gender | 0.0005 | 0.0004 | 0.606 |
| Birth weight | 0.0008 | 0.0000 | 0.004 |
| First birth | 0.0001 | 0.0003 | 0.252 |
| Average household income | 0.0071 | 0.0069 | 0.896 |
| Average density | 0.0070 | 0.0065 | 0.589 |
| Average percent 0–15 year-old | 0.0071 | 0.0069 | 0.792 |
| **C. Time effects** | | | |
| Month | | | |
| February | 0.078 | 0.080 | 0.473 |
| March | 0.086 | 0.085 | 0.783 |
| April | 0.090 | 0.080 | 0.037 |
| May | 0.091 | 0.082 | 0.013 |
| June | 0.091 | 0.083 | 0.023 |
| July | 0.091 | 0.089 | 0.374 |
| August | 0.082 | 0.088 | 0.014 |
| September | 0.075 | 0.083 | 0.024 |
| October | 0.080 | 0.082 | 0.569 |
| November | 0.073 | 0.077 | 0.238 |
| December | 0.077 | 0.081 | 0.291 |
| Day of the week | | | |
| Monday | 0.148 | 0.147 | 0.774 |
| Tuesday | 0.157 | 0.148 | 0.013 |
| Wednesday | 0.141 | 0.141 | 0.960 |
| Thursday | 0.141 | 0.146 | 0.363 |
| Friday | 0.146 | 0.141 | 0.308 |
| Saturday | 0.132 | 0.137 | 0.446 |
| Number of observations | 13,749 | 121,630 | |

Notes: Each cell represents the mean of the corresponding variable in the row after controlling for gestational age. The last column presents the p-value for differences in means clustered at the gestational day level.

Table A3: Robustness to estimation strategy, OB/GYN supervision from the onset of labor

| | 7-day mortality (per 1,000) (1) | 28-day mortality (per 1,000) (2) | Low Apgar score (per cent) (3) |
|---|---|---|---|
| **A. Including controls** | 2.546 | 1.482 | 0.139 |
| | (2.961) | (3.259) | (0.676) |
| Observations | 150,471 | 150,471 | 150,269 |
| **B. Donut sample (excluding days 258 and 259)** | 1.911 | 0.528 | 0.639 |
| | (3.159) | (3.727) | (0.722) |
| Observations | 144,520 | 144,520 | 144,327 |
| **C. Sensitivity to polynomial in gestational age** | | | |
| **Second degree polynomial** | 2.829 | 6.301 | −0.368 |
| | (4.580) | (4.215) | (1.144) |
| Observations | 150,471 | 150,471 | 150,269 |
| **Third degree polynomial** | 9.309 | 14.782*** | −2.898* |
| | (6.856) | (5.176) | (1.482) |
| Observations | 150,471 | 150,471 | 150,269 |
| **D. Sensitivity to bandwidth choice** | | | |
| **7-day bandwidth** | 4.257 | 6.731 | −0.729 |
| | (4.531) | (4.124) | (1.066) |
| Observations | 49,163 | 49,163 | 49,088 |
| **21-day bandwidth** | 1.580 | 1.190 | 0.819 |
| | (2.852) | (3.334) | (0.640) |
| Observations | 369,740 | 369,740 | 369,266 |
| **E. Triangular kernel** | −0.794 | 0.060 | 1.131 |
| | (3.961) | (4.487) | (0.938) |
| Observations | 150,471 | 150,471 | 150,269 |
| **F. Non-linear specification** | | | |
| **First stage:** $W37$ | 0.196*** | 0.196*** | 0.197*** |
| | (0.005) | (0.005) | (0.005) |
| **Reduced form:** $W37$ | 0.562 | 0.458 | 0.068 |
| | (0.776) | (0.820) | (0.148) |
| **IV:** $OB/GYN$ | 2.868 | 2.338 | 0.347 |
| | (3.970) | (4.194) | (0.758) |
| Observations | 150,471 | 150,471 | 150,269 |

Notes: Each cell represents a different regression. Unless otherwise indicated, all specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS; sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Robust standard errors clustered at the gestational day level, except for Panel F which reports bootstrapped standard errors from 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Robustness to sample definition, OB/GYN supervision at the onset of labor

| | 7-day mortality (per 1,000) (1) | 28-day mortality (per 1,000) (2) | Low Apgar score (per cent) (3) |
|---|---|---|---|
| **A. Alternative definitions of the running variable** | | | |
| **Gestational age at birth minus 6 hours** | 4.169 | 3.572 | 0.324 |
| | (3.297) | (3.257) | (0.912) |
| Observations | 155,916 | 155,916 | 155,707 |
| **Gestational age at birth minus 12 hours** | 4.260 | 3.496 | 0.497 |
| | (3.616) | (3.677) | (0.895) |
| Observations | 161,978 | 161,978 | 161,764 |
| **B. Inclusion of referrals to OB/GYN prior to the onset of labor** | | | |
| **Same day as delivery** | 5.108 | 2.599 | 0.855 |
| | (4.502) | (5.240) | (0.984) |
| Observations | 145,549 | 145,549 | 145,351 |
| **All referrals for prematurity** | 2.038 | 1.214 | 0.133 |
| | (2.604) | (3.025) | (0.578) |
| Observations | 151,351 | 151,351 | 151,148 |
| **C. Including observations with missing referral dates** | | | |
| **Coded as supervised by recorded attendant** | 3.653 | 5.653* | 0.367 |
| | (2.624) | (2.952) | (0.712) |
| **All coded as supervised by midwife** | 5.797 | 8.971* | 0.582 |
| | (4.287) | (4.855) | (1.141) |
| **All coded as supervised by OB/GYN** | −5.535 | −8.565* | −0.557 |
| | (4.374) | (4.715) | (1.096) |
| Observations | 226,815 | 226,815 | 226,514 |

Notes: Each cell represents a different regression. Unless otherwise indicated, all specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS; sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Robust standard errors clustered at the gestational day level, except for Panel F which reports bootstrapped standard errors from 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Robustness to estimation strategy, OB/GYN supervision at any point during delivery

| | 7-day mortality (per 1,000) (1) | 28-day mortality (per 1,000) (2) | Low Apgar score (per cent) (3) |
|---|---|---|---|
| **A. Including controls** | 1.091 | 0.635 | 0.060 |
| | (1.287) | (1.393) | (0.290) |
| Observations | 150,471 | 150,471 | 150,269 |
| **B. Donut sample (excluding days 258 and 259)** | 0.823 | 0.228 | 0.276 |
| | (1.382) | (1.602) | (0.311) |
| Observations | 144,520 | 144,520 | 144,327 |
| **C. Sensitivity to polynomial in gestational age** | | | |
| **Second degree polynomial** | 1.113 | 2.478 | −0.145 |
| | (1.847) | (1.666) | (0.457) |
| Observations | 150,471 | 150,471 | 150,269 |
| **Third degree polynomial** | 3.407 | 5.411** | −1.064* |
| | (2.794) | (2.203) | (0.589) |
| Observations | 150,471 | 150,471 | 150,269 |
| **D. Sensitivity to bandwidth choice** | | | |
| **7-day bandwidth** | 1.677 | 2.652 | −0.288 |
| | (1.859) | (1.661) | (0.433) |
| Observations | 49,163 | 49,163 | 49,088 |
| **21-day bandwidth** | 0.671 | 0.505 | 0.348 |
| | (1.215) | (1.411) | (0.269) |
| Observations | 369,740 | 369,740 | 369,266 |
| **E. Triangular kernel** | −0.322 | 0.024 | 0.459 |
| | (1.591) | (1.818) | (0.375) |
| Observations | 150,471 | 150,471 | 150,269 |
| **F. Non-linear specification** | | | |
| **First stage:** $W37$ | 0.456*** | 0.456*** | 0.457*** |
| | (0.006) | (0.006) | (0.006) |
| **Reduced form:** $W37$ | 0.562 | 0.458 | 0.068 |
| | (0.776) | (0.820) | (0.148) |
| **IV:** $OB/GYN$ | 1.232 | 1.004 | 0.149 |
| | (1.702) | (1.800) | (0.325) |
| Observations | 150,471 | 150,471 | 150,269 |

Notes: Each cell represents a different regression. Unless otherwise indicated, all specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS; sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Robust standard errors clustered at the gestational day level, except for Panel F which reports bootstrapped standard errors from 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Robustness to sample definition, OB/GYN supervision at any point during delivery

| | 7-day mortality (per 1,000) (1) | 28-day mortality (per 1,000) (2) | Low Apgar score (per cent) (3) |
|---|---|---|---|
| **A. Alternative definitions of the running variable** | | | |
| **Gestational age at birth minus 6 hours** | 1.773 | 1.519 | 0.138 |
| | (1.429) | (1.384) | (0.388) |
| Observations | 155,916 | 155,916 | 155,707 |
| **Gestational age at birth minus 12 hours** | 1.826 | 1.498 | 0.213 |
| | (1.556) | (1.564) | (0.387) |
| Observations | 161,978 | 161,978 | 161,764 |
| **B. Inclusion of referrals to OB/GYN prior to the onset of labor** | | | |
| **Same day as delivery** | 1.576 | 0.802 | 0.264 |
| | (1.405) | (1.611) | (0.302) |
| Observations | 145,549 | 145,549 | 145,351 |
| **All referrals for prematurity** | 0.960 | 0.572 | 0.063 |
| | (1.245) | (1.423) | (0.272) |
| Observations | 151,351 | 151,351 | 151,148 |
| **C. Including observations with missing referral dates** | | | |
| **Coded as supervised by recorded attendant** | 2.039 | 3.155* | 0.205 |
| | (1.515) | (1.657) | (0.400) |
| **All coded as supervised by midwife** | 2.616 | 4.048* | 0.263 |
| | (1.988) | (2.178) | (0.516) |
| **All coded as supervised by OB/GYN** | 34.331 | 53.122** | 3.474 |
| | (24.427) | (25.245) | (6.737) |
| Observations | 226,815 | 226,815 | 226,514 |

Notes: Each cell represents a different regression. Unless otherwise indicated, all specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS; sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Robust standard errors clustered at the gestational day level, except for Panel F which reports bootstrapped standard errors from 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Heterogeneous effects, OB/GYN supervision at the onset of labor

| | First stage | 7-day mortality (per 1,000) | 28-day mortality (per 1,000) | Low Apgar score (per cent) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Baseline results** | 0.198*** | 3.631 | 2.725 | 0.352 |
| | (0.017) | (2.940) | (3.368) | (0.644) |
| Observations | | 150,471 | 150,471 | 150,269 |
| Average health outcome | | 1.447 | 1.617 | 0.710 |
| **A. Mother: Non-Dutch** | 0.184*** | −8.577 | −7.031 | −0.528 |
| | (0.012) | (10.353) | (12.993) | (1.155) |
| Observations | | 31,293 | 31,293 | 31,259 |
| Average health outcome | | 1.694 | 1.911 | 0.888 |
| **B. Mother's age** | | | | |
| **Younger than median (30 years)** | 0.204*** | 3.075 | 3.468 | 0.269 |
| | (0.019) | (4.167) | (4.924) | (0.981) |
| Observations | | 82,099 | 82,099 | 81,991 |
| Average health outcome | | 1.381 | 1.580 | 0.715 |
| **Older than median (30 years)** | 0.191*** | 4.174 | 1.464 | 0.454 |
| | (0.017) | (5.885) | (5.508) | (1.504) |
| Observations | | 68,372 | 68,372 | 68,278 |
| Average health outcome | | 1.525 | 1.660 | 0.703 |
| **C. Infant: Low birth weight (1,500-2,500g)** | 0.164*** | 11.512 | 10.422 | 2.124 |
| | (0.018) | (19.788) | (19.853) | (3.649) |
| Observations | | 9,440 | 9,440 | 9,424 |
| Average health outcome | | 6.917 | 7.350 | 2.383 |
| **D. Infant: Parity** | | | | |
| **First birth** | 0.210*** | 3.167 | 3.224 | 0.127 |
| | (0.017) | (2.274) | (2.944) | (0.761) |
| Observations | | 79,541 | 79,541 | 79,431 |
| Average health outcome | | 1.425 | 1.623 | 0.865 |
| **Higher-order birth** | 0.178*** | 4.131 | 0.760 | 0.620 |
| | (0.018) | (7.313) | (7.565) | (1.486) |
| Observations | | 70,930 | 70,930 | 70,838 |
| Average health outcome | | 1.470 | 1.610 | 0.549 |

Table A7: Heterogeneous effects, OB/GYN supervision at the onset of labor (cont'd)

| | First stage | 7-day mortality (per 1,000) | 28-day mortality (per 1,000) | Low Apgar score (per cent) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **E. Average household income** | | | | |
| **First quartile** | 0.189*** | −19.416*** | −19.494*** | −2.243* |
| | (0.021) | (5.375) | (6.570) | (1.213) |
| Observations | | 37,716 | 37,716 | 37,659 |
| Average health outcome | | 1.746 | 1.868 | 0.899 |
| **Second quartile** | 0.192*** | 11.573 | 8.485 | 1.249 |
| | (0.019) | (7.450) | (9.450) | (1.911) |
| Observations | | 37,536 | 37,536 | 37,476 |
| Average health outcome | | 1.145 | 1.300 | 0.648 |
| **Third quartile** | 0.206*** | 2.566 | 0.614 | 0.835 |
| | (0.015) | (5.162) | (5.325) | (1.185) |
| Observations | | 37,633 | 37,633 | 37,590 |
| Average health outcome | | 1.355 | 1.602 | 0.604 |
| **Fourth quartile** | 0.205*** | 18.691** | 20.095*** | 1.502 |
| | (0.017) | (9.102) | (7.155) | (0.979) |
| Observations | | 37,586 | 37,586 | 37,544 |
| Average health outcome | | 1.540 | 1.694 | 0.688 |

Notes: Each cell represents a different regression. All specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS; sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Column 1 lists the first stage coefficient of $W37$ in the the mortality sample; columns 2-4 the coefficient of $OB/GYN$ in the structural equation. Robust standard errors clustered at the gestational day level. The average health outcome is for observations to the right of the cutoff (gestational age greater than 258 days). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Heterogeneous effects, OB/GYN supervision at any point during delivery

| | First stage | 7-day mortality (per 1,000) | 28-day mortality (per 1,000) | Low Apgar score (per cent) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Baseline results** | 0.463*** | 1.555 | 1.167 | 0.151 |
| | (0.036) | (1.291) | (1.438) | (0.275) |
| Observations | | 150,471 | 150,471 | 150,269 |
| Average health outcome | | 1.447 | 1.617 | 0.710 |
| **A. Mother: Non-Dutch** | 0.463*** | −3.416 | −2.800 | −0.211 |
| | (0.027) | (3.952) | (5.065) | (0.461) |
| Observations | | 31,293 | 31,293 | 31,259 |
| Average health outcome | | 1.694 | 1.911 | 0.888 |
| **B. Mother's age** | | | | |
| **Younger than median (30 years)** | 0.472*** | 1.329 | 1.499 | 0.116 |
| | (0.034) | (1.780) | (2.096) | (0.424) |
| Observations | | 82,099 | 82,099 | 81,991 |
| Average health outcome | | 1.381 | 1.580 | 0.715 |
| **Older than median (30 years)** | 0.452*** | 1.762 | 0.618 | 0.192 |
| | (0.039) | (2.520) | (2.320) | (0.627) |
| Observations | | 68,372 | 68,372 | 68,278 |
| Average health outcome | | 1.525 | 1.660 | 0.703 |
| **C. Infant: Low birth weight (1,500-2,500g)** | 0.434*** | 4.356 | 3.944 | 0.807 |
| | (0.031) | (7.377) | (7.347) | (1.363) |
| Observations | | 9,440 | 9,440 | 9,424 |
| Average health outcome | | 6.917 | 7.350 | 2.383 |
| **D. Infant: Parity** | | | | |
| **First birth** | 0.447*** | 1.488 | 1.515 | 0.060 |
| | (0.028) | (1.063) | (1.352) | (0.358) |
| Observations | | 79,541 | 79,541 | 79,431 |
| Average health outcome | | 1.425 | 1.623 | 0.865 |
| **Higher-order birth** | 0.489*** | 1.506 | 0.277 | 0.227 |
| | (0.051) | (2.725) | (2.763) | (0.536) |
| Observations | | 70,930 | 70,930 | 70,838 |
| Average health outcome | | 1.470 | 1.610 | 0.549 |

Table A9: Heterogeneous effects, OB/GYN supervision at any point during delivery (cont'd)

| | First stage | 7-day mortality (per 1,000) | 28-day mortality (per 1,000) | Low Apgar score (per cent) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **E. Average household income** | | | | |
| **First quartile** | 0.441*** | −8.334*** | −8.368*** | −0.962* |
| | (0.034) | (2.275) | (2.961) | (0.531) |
| Observations | | 37,716 | 37,716 | 37,659 |
| Average health outcome | | 1.746 | 1.868 | 0.899 |
| **Second quartile** | 0.464*** | 4.795 | 3.516 | 0.518 |
| | (0.042) | (2.928) | (3.774) | (0.780) |
| Observations | | 37,536 | 37,536 | 37,476 |
| Average health outcome | | 1.145 | 1.300 | 0.648 |
| **Third quartile** | 0.495*** | 1.071 | 0.256 | 0.349 |
| | (0.034) | (2.186) | (2.230) | (0.498) |
| Observations | | 37,633 | 37,633 | 37,590 |
| Average health outcome | | 1.355 | 1.602 | 0.604 |
| **Fourth quartile** | 0.454*** | 8.462* | 9.098** | 0.681 |
| | (0.037) | (4.283) | (3.324) | (0.440) |
| Observations | | 37,586 | 37,586 | 37,544 |
| Average health outcome | | 1.540 | 1.694 | 0.688 |

Notes: Each cell represents a different regression. All specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS; sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Column 1 lists the first stage coefficient of $W37$ in the the mortality sample; columns 2-4 the coefficient of $OB/GYN$ in the structural equation. Robust standard errors clustered at the gestational day level. The average health outcome is for observations to the right of the cutoff (gestational age greater than 258 days). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A10: Complier characteristics, relative likelihoods

| | Relative likelihood | | Observations |
|---|---|---|---|
| | Obstetrician supervision at the onset of labor (1) | Obstetrician supervision at any point during delivery (2) | (3) |
| **A. Maternal characteristics** | | | |
| Ethnicity: Dutch | 1.016 | 1.000 | 119,178 |
| Age | | | |
| Less than 35 | 1.022 | 1.013 | 133,736 |
| Less than median (30 years) | 1.030 | 1.020 | 82,099 |
| Intended home birth | 1.006 | 0.943 | 55,981 |
| **B. Infant characteristics** | | | |
| Birth weight | | | |
| Less than median (3,160g) | 1.003 | 1.023 | 75,980 |
| Low birth weight (1,500–2,500g) | 0.829 | 0.938 | 9,440 |
| First birth | 1.060 | 0.966 | 79,541 |
| **C. Postal code characteristics** | | | |
| Average household income | | | |
| Below median (1,921 euros) | 0.962 | 0.976 | 75,252 |
| First quartile | 0.954 | 0.952 | 37,716 |
| Second quartile | 0.969 | 1.001 | 37,536 |
| Third quartile | 1.041 | 1.068 | 37,633 |
| Fourth quartile | 1.036 | 0.980 | 37,586 |
| Share of compliers | 0.198 | 0.463 | 150,471 |

Notes: Column 1 shows the likelihood that compliers have the characteristic indicated in the row relative to the entire analysis sample within our chosen bandwidth. Column 2 shows the number of observations with that characteristic in the analysis sample. The last row gives the share of compliers (based on the first-stage in the mortality sample) and total number of observations in the analysis sample.

Table A11: NICU admissions (per cent)

| | All | Average Household Income | | | |
|---|---|---|---|---|---|
| | | 1st quartile | 2nd quartile | 3rd quartile | 4th quartile |
| | (1) | (2) | (3) | (4) | (5) |
| **A. OB/GYN supervision from the onset of labor** | | | | | |
| **First stage:** $W37$ | 0.198*** | 0.189*** | 0.192*** | 0.206*** | 0.205*** |
| | (0.017) | (0.021) | (0.019) | (0.015) | (0.017) |
| **Reduced form:** $W37$ | 14.774*** | 13.378*** | 14.424*** | 16.467*** | 14.857*** |
| | (1.198) | (1.974) | (1.169) | (0.819) | (1.869) |
| **Instrumental variable:** $OB/GYN$ | 74.503*** | 70.680*** | 75.059*** | 79.783*** | 72.331*** |
| | (4.430) | (5.683) | (3.922) | (4.281) | (9.884) |
| Average of health outcome after gestational day 258 | 7.927 | 7.857 | 8.732 | 8.183 | 6.939 |
| Observations | 150,471 | 37,716 | 37,536 | 37,633 | 37,586 |
| **B. OB/GYN supervision at any point during delivery** | | | | | |
| **First stage:** $W37$ | 0.463*** | 0.441*** | 0.464*** | 0.495*** | 0.454*** |
| | (0.036) | (0.034) | (0.042) | (0.034) | (0.037) |
| **Reduced form:** $W37$ | 14.774*** | 13.378*** | 14.424*** | 16.467*** | 14.857*** |
| | (1.198) | (1.974) | (1.169) | (0.819) | (1.869) |
| **Instrumental variable:** $OB/GYN$ | 31.901*** | 30.339*** | 31.101*** | 33.288*** | 32.747*** |
| | (0.831) | (2.529) | (1.615) | (1.776) | (3.397) |
| Average of health outcome after gestational day 258 | 7.927 | 7.857 | 8.732 | 8.183 | 6.939 |
| Observations | 150,471 | 37,716 | 37,536 | 37,633 | 37,586 |

Notes: Each cell represents a different regression. All specifications include a first-degree polynomial in normalized gestational age and its interaction with the week-37 indicator and are estimated by OLS. Sample restricted to observations with gestational age within a 14-day bandwidth around day 258. Robust standard errors clustered at the gestational day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$